PTO/SB/17 (12-04)
Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2006

Effective 10/01/2003. Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) **180.00**

Complete if Known	
Application Number	10/006,818
Filing Date	December 6, 2001
First Named Inventor	Kevin P. Baker
Examiner Name	Fozia M. Hamud
Art Unit	1647
Attorney Docket No.	39780-2830 P1C4

METHOD OF PAYMENT *(check all that apply)*

☐ Check ☐ Credit card ☐ Money Order ☐ Other ☐ None

☑ Deposit Account:

Deposit Account Number: **08-1641(39780-2830P1C4)**

Deposit Account Name: **HELLER EHRMAN, LLP**

The Director is authorized to: *(check all that apply)*

☑ Charge fee(s) indicated below ☑ Credit any overpayments

☑ Charge any additional fee(s) or any underpayment of fee(s)

☐ Charge fee(s) indicated below, **except for the filing fee** to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	300	2001	150	Utility filing fee	
1002	200	2002	100	Design filing fee	
1003	200	2003	100	Plant filing fee	
1004	300	2004	150	Reissue filing fee	
1005	200	2005	100	Provisional filing fee	

SUBTOTAL (1) ($)

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

		Extra Claims		Fee from below	Fee Paid
Total Claims	-20** =		X		=
Independent Claims	- 3** =		X		=
Multiple Dependent					=

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	50	2202	25	Claims in excess of 20
1201	200	2201	100	Independent claims in excess of 3
1203	360	2203	180	Multiple dependent claim, if not paid
1204	200	2204	100	** Reissue independent claims over original patent
1205	50	2205	25	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($)

***or number previously paid, if greater; For Reissues, see above*

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	120	2251	60	Extension for reply within first month	
1252	450	2252	225	Extension for reply within second month	
1253	1,020	2253	510	Extension for reply within third month	
1254	1,590	2254	795	Extension for reply within fourth month	
1255	2,160	2255	1,080	Extension for reply within fifth month	
1401	500	2401	250	Notice of Appeal	
1402	500	2402	250	Filing a brief in support of an appeal	
1403	1,000	2403	500	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	500	2452	250	Petition to revive - unavoidable	
1453	1,500	2453	750	Petition to revive - unintentional	
1501	1,400	2501	700	Utility issue fee (or reissue)	
1502	800	2502	400	Design issue fee	
1503	1,100	2503	550	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	180.00
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	790	2809	395	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	790	2810	395	For each additional invention to be examined (37 CFR 1.129(b))	
1801	790	2801	395	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($) **180.00**

SUBMITTED BY

(Complete *(if applicable)*)

Name (Print/Type)	Barrie D. Greene	Registration No. (Attorney/Agent)	46,740	Telephone	(650) 324-7000
Signature				Date	June 2, 2006

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

This collection of information is required by 37 CFR 1.17 and 1.27. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	Examiner: Hamud, Fozia M.
Kevin P. BAKER, et al.	Art Unit: 1647
Application Serial No. 10/006,818	Confirmation No: 1321
Filed: December 6, 2001	Attorney's Docket No. 39780-2830 P1C4
For: **SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME**	**Customer No. 35489**

EXPRESS MAIL LABEL NO. EV 765 973 322 US
DATE MAILED: JUNE 2, 2006

REQUEST FOR REOPENING OF PROSECUTION AND RESPONSE UNDER 37 C.F.R. §1.111

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Dear Sir:

On January 3, 2005, the Examiner made a final rejection to pending Claims 28-32. A Notice of Appeal was filed on May 27, 2005, and Appellants' Appeal Brief was filed on July 26, 2005. A Notice of Non-Compliant Appeal Brief was mailed October 31, 2005, and a revised Appeal Brief was filed November 22, 2005.

An Examiner's Answer was mailed on April 3, 2006, which contains new grounds of rejection. Applicants were granted two months from the mailing date of the Examiner's Answer to request that the prosecution be reopened. In response, Applicants request that the prosecution be reopened under 37 C.F.R.§41.39. In addition, Applicants submit herein a Response under 37 C.F.R. §1.111. The Response and the Request are timely filed within the two-month period for response set by the Examiner's Answer.

This Response is concurrently filed with the submission of a new Declaration under 37 C.F.R. §1.132 by Dr. Paul Polakis, with attached Exhibits A and B. Also filed herewith is an Information Disclosure Statement providing the article by Beer *et al.* Applicants respectfully

request that the information listed in the Information Disclosure Statement be considered by the Examiner and be made of record in the above-identified application.

Remarks/Arguments begin on page 3 of this paper.

REMARKS/ARGUMENTS

Claim Rejections Under 35 U.S.C. §102

Applicants acknowledge the Examiner's statement that the rejection of Claims 28-32 under 35 U.S.C. §102(a) as allegedly being anticipated by Botstein *et al.* (WO 2000053751) is withdrawn, because the instant application is entitled to an effective filing date of February 18, 2000.

Claim Rejections Under 35 U.S.C. §101 and §112, First Paragraph, Enablement

Claims 28-32 are rejected under 35 U.S.C. §101 as allegedly lacking either a specific and substantial asserted utility or a well-established utility. Claims 28-32 are further rejected under 35 U.S.C. §112, first paragraph, as allegedly lacking enablement "since the claimed invention is not supported by either a credible, specific and substantial utility or a well established utility ... one skilled in the art clearly would not know how to use the claimed invention." (Page 10 of the Examiner's Answer). In her Answer, the Examiner acknowledges that the gene encoding PRO1293 is amplified in certain human lung and colon cancers. However, the Examiner argues that the gene amplification data do not provide utility or enablement for the PRO1293 polypeptide or the claimed antibodies that bind it. The Examiner makes the following arguments in support of these conclusions:

(1) the PRO1293 gene was amplified in only 3 of the disclosed lung and colon tumors and tumor cell lines;

(2) the gene amplification assay used a pooled normal blood control instead of a matched tissue control, which is allegedly the standard in the art;

(3) an at least 2-fold amplification of DNA in tumors is allegedly not considered by the literature to be significant;

(4) the literature allegedly shows that there is no correlation between gene amplification and increased mRNA expression;

(5) the art allegedly shows that there is no correlation between mRNA levels and polypeptide levels in tumors or in normal tissues; and

(6) the Polakis Declaration does not provide support for Applicant's assertions of utility because it does not provide data so that the Examiner can independently draw conclusions.

Applicants disagree with each of the Examiner's arguments for the reasons detailed below.

The Examiner asserts that the significance of the gene amplification data "can be questioned since 49 out of 52 tested tumor samples did not show an amplification of the gene encoding PRO1293." (Pages 12-13 of the Examiner's Answer). Applicants emphasize that they have shown significant DNA amplification in three of the lung and colon tumor samples in Table 8, Example 143 of the instant specification. The fact that not all lung and colon tumors tested positive in this study does not make the gene amplification data less significant. As any skilled artisan in the field of oncology would easily appreciate, not all tumor markers are generally associated with every tumor, or even with most tumors. For example, the article by Hanna and Mornin (submitted with the Response filed August 19, 2004), discloses that the known breast cancer marker HER-2/neu is "amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma" (page 1, col. 1). In fact, some tumor markers are useful for identifying rare malignancies. That is, the association of the tumor marker with a particular type of tumor lesion may be rare, or, the occurrence of that particular kind of tumor lesion itself may be rare. In either event, even these rare tumor markers which do not give a positive hit for most common tumors, have great value in tumor diagnosis, and consequently, in tumor prognosis. The skilled artisan would certainly know that such tumor markers are useful for better classification of tumors. Therefore, whether the PRO1293 gene is amplified in three lung and colon tumors or in all lung and colon tumors is not relevant to its identification as a tumor marker, or its patentable utility. Rather, the fact that the amplification data for PRO1293 is considered significant is what lends support to its usefulness as a tumor marker.

The Examiner further asserts that the gene amplification data are not persuasive because "the control used was not a matched non-tumor lung sample but rather was a pooled DNA sample from the blood of healthy subjects. The art uses matched tissue samples (see Pennica *et al.*)." (Page 13 of the Examiner's Answer).

Applicants respectfully submit that the negative control taught in the specification was known in the art at the time of filing, and accepted as a true negative control as demonstrated by use in peer reviewed publications, including Pennica *et al.* For example, Pennica *et al.* explains that "[t]he relative WISP gene copy number in each colon tumor DNA was compared with **pooled normal DNA** from 10 donors by quantitative PCR" (page 14720, col. 2; emphasis

added). Pennica *et al.* further explain that DNA was isolated from "the pooled blood of 10 normal human donors" (page 14718, col. 1). Thus Pennica *et al.* used the **same** control for their gene amplification experiments as that described in the instant specification.

In further examples, Pitti *et al.* (Exhibit F submitted with the Response filed August 19, 2004), used the same quantitative TaqMan PCR assay described in the specification to study gene amplification in lung and colon cancer of DcR3, a decoy receptor for Fas ligand. As described, Pitti *et al.* analyzed DNA copy number "in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBL) of 10 healthy donors." (Page 701, col. 1; emphasis added). The authors also analyzed mRNA expression of DcR3 in primary tumor tissue sections and found tumor-specific expression, confirming the finding of frequent amplification in tumors, and confirming that the pooled blood sample was a valid negative control for the gene amplification experiments. In Bieche *et al.* (Exhibit G submitted with the Response filed August 19, 2004), the authors used the quantitative TaqMan PCR assay to study gene amplification of *myc*, *ccnd1* and *erbB2* in breast tumors. As their negative control, Bieche *et al.* used normal leukocyte DNA derived from a small subset of the breast cancer patients (page 663). The authors note that "[t]he results of this study are consistent with those reported in the literature" (page 664, col. 2), thus confirming the validity of the negative control. Accordingly, the art demonstrates that pooled normal blood samples are considered to be a valid negative control for gene amplification experiments of the type described in the specification.

The Examiner asserts that "[t]he specification merely demonstrates that the PRO1293 genomic DNA was amplified in some cancers, to a minor degree (about 2-5 fold) relative to normal blood DNA." (Page 13 of the Examiner's Answer).

Applicants respectfully submit that the Examiner seems to be applying a heightened utility standard in this instance, which is legally incorrect. Applicants have shown that the gene encoding PRO1293 demonstrated significant amplification, from 2.19 to 5.03 fold, in three lung and colon tumors. As explained in the Declaration of Dr. Audrey Goddard (submitted with the Response filed August 19, 2004):

> It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample **is significant** and useful in that the detected increase in gene

> copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. (Emphasis added).

By referring to the 2.19-fold to 5.03-fold amplification of the PRO1293 gene in lung and colon tumors as "minor" the Examiner appears to ignore the teachings within an expert's declaration without any basis, or without presenting any evidence to the contrary. Applicants respectfully draw the Examiner's attention to the Utility Examination Guidelines (Part IIB, 66 Fed. Reg. 1098 (2001)) which state that:

> Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered.

Thus, barring evidence to the contrary, Applicants maintain that the 2.17 to 5.03-fold amplification disclosed for the PRO1293 gene is significant and forms the basis for the utility claimed herein.

The Examiner asserts that the Goddard Declaration is not convincing, because the six references submitted with the Declaration allegedly do not "appear to indicate that an approximately 2-5 fold amplification of genomic DNA is significant in tumors." (Page 13 of the Examiner's Answer). Applicants respectfully submit that this statement is scientifically and factually inaccurate. The three references which discuss applications of the PCR-based gene amplification determination technique to studies of specific genes make clear that values of at least 2-fold in the assay of Example 143 are considered to meet the threshold for significant amplification in tumors.

In Pennica *et al*, for example, the authors concluded that WISP-1 was aberrantly expressed in human colon tumors based upon an observed amplification of at least 2-fold in about 60% of the tumors tested (page 14720, col. 2). Similarly, in Pitti *et al.*, the authors concluded that DcR3 was amplified in lung and colon tumors based upon an observed amplification ranging from 2 to 18-fold in about half of the tumors tested. In Bieche *et al.*, the authors explicitly state that "**values of 2 or more** were considered to represent gene amplification in tumor DNA" (page 664, col. 1; emphasis added). Thus the art is clear that an observed amplification of at least 2-fold in the assay of Example 143 is considered to be

indicative of significant amplification in tumors, sufficient to demonstrate that amplification of the gene is associated with tumors.

Accordingly, Applicants submit that based on the general knowledge in the art at the time the invention was made and the teachings in the specification, the specification provides clear guidance as to how to interpret and use the data relating to PRO1293 expression and that the PRO1293 polypeptide and the claimed antibodies that bind it have utility in the diagnosis of cancer.

A *prima facie* case of lack of utility has not been established

The Examiner has asserted that the disclosed gene amplification data does not establish a patentable utility for the PRO1293 polypeptides because allegedly "it does not necessarily follow that an increase in gene copy (DNA) number results in increased gene expression (mRNA) and increased protein expression such that the polypeptide of SEQ ID NO:77, or variants of the polypeptide of SEQ ID NO:77, would be useful diagnostically." (Pages 5-6 of the Examiner's Answer). In support of the assertion that gene amplification is not correlated with increased mRNA expression, the Examiner refers to Pennica *et al.*, as well as a newly cited reference by Konopka *et al.* (Page 6 of the Examiner's Answer). The Examiner further asserts that "[e]ven if increased mRNA levels could be established for PRO1293, it does not follow that polypeptide levels would also be amplified," referring to Hu *et al.* and a newly cited reference by Chen *et al.* for support. (Pages 6-7 of the Examiner's Answer). Finally, the Examiner asserts that "[t]he art also shows that mRNA (transcript) levels do not correlate with polypeptide levels in normal tissues, citing five new references by Haynes *et al.*, Gygi *et al.*, Lian *et al.*, Fessler *et al.* and Greenbaum *et al.* (Pages 7-9 of the Examiner's Answer).

As a preliminary matter, Applicants respectfully submit that it is not a legal requirement to establish that gene amplification necessarily results in increased expression at the mRNA and polypeptide levels, or that protein levels can be "accurately predicted." As discussed in Applicants' Appeal Brief, the evidentiary standard to be used throughout *ex parte* examination of a patent application is a preponderance of the totality of the evidence under consideration. Accordingly, Applicants submit that in order to overcome the presumption of truth that an assertion of utility by the Applicant enjoys, the Examiner must establish that **it is more likely than not** that one of ordinary skill in the art would doubt the truth of the statement of utility.

Therefore, it is not legally required that there be a "necessary" correlation between the data presented and the claimed subject matter. The law requires only that one skilled in the art should accept that such a correlation is **more likely than not to exist**. Applicants respectfully submit that when the proper evidentiary standard is applied, a correlation must be acknowledged.

Pennica et al. and Konopka et al.

In support of the assertion that gene amplification is not correlated with increased mRNA expression, the Examiner refers to Pennica *et al.*, as well as a newly cited reference by Konopka *et al.* (Page 6 of the Examiner's Answer). In particular, the Examiner cites the abstract of Pennica *et al.* for its disclosure that "WISP-1 gene amplification and overexpression in human colon tumors showed a correlation between DNA amplification and over-expression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with expression in normal colonic mucosa from the same patient." From this, the Examiner correctly concludes that increased copy number does not *necessarily* result in increased polypeptide expression. The standard, however, is not absolute certainty.

As noted even in Pennica *et al.*, "[a]n analysis of *WISP*-1 gene amplification and expression in human colon tumors *showed a correlation between DNA amplification and over-expression...*" (Pennica *et al.*, page14722, left column, first full paragraph, emphasis added). Thus the findings of Pennica *et al.* with respect to WISP-1 support Applicants' arguments. In the case of WISP-3, the authors report that there was no change in the DNA copy number, but there was a change in mRNA levels. This apparent lack of correlation between DNA and mRNA levels is not contrary to Applicants' assertion that a change in DNA copy number generally leads to a change in mRNA level. Applicants are not attempting to predict the DNA copy number based on changes in mRNA level, and Applicants have not asserted that the only means for changing the level of mRNA is to change the DNA copy number. Therefore a change in mRNA without a change in DNA copy number is not contrary to Applicants' assertions.

The fact that the single WISP-2 gene did not show the expected correlation of gene amplification with the level of mRNA/protein expression does not establish that it is more likely than not, in general, that such correlation does not exist. The Examiner has not shown whether

the lack or correlation observed for the WISP-2 gene is typical, or is merely a discrepancy, an exception to the rule of correlation. Indeed, the working hypothesis among those skilled in the art is that, if a gene is amplified in cancer, the encoded protein is likely to be expressed at an elevated level, as was demonstrated for WISP-1.

Accordingly, Applicants respectfully submit that Pennica *et al.* teaches nothing conclusive regarding the absence of correlation between amplification of a gene and over-expression of the encoded WISP polypeptide. More importantly, the teaching of Pennica *et al.* is specific to *WISP* genes. Pennica *et al.* has no teaching whatsoever about the correlation of gene amplification and protein expression in general.

The Examiner argues that Pennica *et al* is relevant even though it is limited to only one gene family because it is "shows a lack of correlation between gene amplification and gene product overexpression" and because the instant case also concerns a single gene. (Page 15 of the Examiner' Answer). Applicants respectfully disagree. The test is whether it is more likely than not that gene amplification results in overexpression of the corresponding mRNA and protein. In order to meet that standard, the Examiner must provide evidence that it is more likely than not that gene amplification does not result in mRNA or protein overexpression. Providing the single example of the WISP-2 gene does not suffice to meet this burden.

Applicants next respectfully submit that, contrary to the PTO's assertions, Konopka *et al.* **supports** Applicants' position that mRNA levels correlate with protein levels. Konopka *et al.* states that "the 8-kb mRNA that encodes $P210^{c\text{-}abl}$ was detected at a 10-fold higher level in SK-CML7bt-333 (Fig. 3A, +) than in SK-CML16Bt-1 (B, +), which **correlated** with the relative level of $P210^{c\text{-}abl}$ detected in each cell line. Analysis of additional cell lines demonstrated that the level of 8-kb mRNA **directly correlated** with the level of $P210^{c\text{-}abl}$ (Table 1)" (page 4050, col. 2, emphasis added).

Nor does Konopka *et al.* support the PTO's position that DNA amplification is not correlated with mRNA or protein overexpression. Konopka *et al.* show only that, of the cell lines known to have increased abl protein expression, only one had amplification of the abl gene (page 4051, col. 1). This result proves only that increased mRNA and protein expression levels can result from causes other than gene amplification. Konopka *et al.* do not demonstrate that when gene amplification does occur, it does not result in increased mRNA and protein

expression levels, particularly given that the cell line with amplification of the abl gene did show increased abl mRNA and protein expression levels.

Hu et al. and Chen et al.

In support of the assertion that "[e]ven if increased mRNA levels could be established for PRO1293, it does not follow that polypeptide levels would also be amplified," the Examiner refers to Hu *et al.* and a newly cited reference by Chen *et al.* for support. (Pages 6-7 of the Examiner's Answer). In particular, the Examiner cites Hu *et al.* to the effect that genes displaying a 5-fold change or less in mRNA expression in tumors compared to normal showed no evidence of a correlation between altered gene expression and a known role in the disease. However, among genes with a 10-fold or more change in expression level, there was a strong and significant correlation between expression level and a published role in the disease. (Pages 6-7 of the Examiner's Answer).

Applicants submit that in order to overcome the presumption of truth that an assertion of utility by the Applicant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Accordingly, contrary to the Examiner's assertion, Applicants submit that Hu *et al.* does not conclusively show that it is more likely than not that gene amplification does not result in increased expression at the mRNA and polypeptide levels.

Applicants respectfully point out that the analysis by Hu *et al.* has certain statistical flaws. According to Hu *et al.*, "*different* statistical methods 'were applied to' *estimate* the strength of gene-disease relationships and evaluated the results." (See page 406, left column, emphasis added). Using these different statistical methods, Hu *et al.* "[a]ssessed the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation." (See page 411, left column). It is well known in the art that various statistical methods allow different variables to be manipulated to affect the outcome. For example, the authors admit, "Initial attempts to search the literature using" the list of genes, gene names, gene symbols, and frequently used synonyms, generated by the authors "revealed several sources of false positives and false negatives." (See page 406, right column). The authors further admit that the false positives caused by "duplicative and unrelated meanings for the term" were "difficult to manage." Therefore, in order to minimize such false positives, Hu *et al.* disclose that these terms

"had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes." (See page 406, right column). Hence, Applicants respectfully submit that in order to minimize the false positives and negatives in their analysis, Hu *et al.* manipulated various aspects of the input data.

Applicants further submit that the statistical analysis by Hu *et al.* is not a reliable standard because the frequency of citation only reflects the current research interest in a molecule, not the true biological function of the molecule. Indeed, the authors acknowledge that "[r]elationships established by frequency of co-citation do not necessarily represent a true biological link." (See page 411, right column). One would expect that genes with the greatest change in expression in a disease would be the first targets of research, and therefore have the strongest known relationship to the disease as measured by the number of publications reporting a connection with the disease. The correlation reported in Hu only indicates that the greater the change in expression level, the more likely it is that there is a published or known role for the gene in the disease, as found by their automated literature-mining software. Thus, Hu's results merely reflect a bias in the literature toward studying the most prominent targets, and say nothing regarding the ability of a gene that is 2-fold or more differentially expressed in tumors to serve as a disease marker.

Even assuming that Hu *et al.* provide evidence to support a true relationship, the conclusion in Hu *et al.* only applies to a specific type of breast tumor (estrogen receptor (ER)-positive breast tumor) and can not be generalized as a principle governing microarray study of breast cancer in general, let alone the various other types of cancer genes in general. In fact, even Hu *et al* admit that., "[i]t is likely that this threshold will change depending on the disease as well as the experiment. Interestingly, the observed correlation was only found among ER-positive (breast) tumors not ER-negative tumors." (See page 412, left column). Therefore, based on these findings, the authors add, "[t]his may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently." (See page 412, left column; emphasis added).

Furthermore, Hu *et al.* did not look for a correlation between changes in mRNA and changes in protein levels, and therefore their results are not contrary to Applicants' assertion that

there is a correlation between the two. Applicants are not relying on any "biological role" that the PRO1293 polypeptide has in cancer for its asserted utility. Instead, Applicants are relying on the overexpression of PRO1293 in certain tumors compared to their normal tissue counterparts. Nowhere in Hu does it say that a lack of correlation in their study means that genes with a less than five-fold change in level of expression in cancer cannot serve as a diagnostic marker of cancer.

The Examiner asserts that "Appellant is holding Hu *et al.* to a higher standard than their own specification" for statistical analysis. (Page 17 of the Examiner's Answer). However, Applicants have compared the level of amplification of the PRO1293 gene in normal tissue and lung and colon tumors and have provided information indicating a greater than 2-fold amplification. Applicants are not relying on statistical analysis of information obtained from published literature based on the current research interest of a molecule, and hence the issues regarding statistical analysis of such information do not apply to Applicants' data.

The Examiner further cites a new reference by Chen *et al.* as allegedly disclosing that "only 17% of 165 polypeptide spots or 21% of the genes had a significant correlation between protein and mRNA expression levels" in lung adenocarcinoma samples. (Page 6 of the Examiner's Answer).

First, Applicants note that proteins selected for study by Chen *et al.* were those detectable by staining of 2D gels. As noted in, for example, Haynes *et al.*, cited by the Examiner in the Examiner's Answer, there are problems with selecting proteins detectable by 2D gels. "It is apparent that without prior enrichment only a relatively small and highly selected population of long-lived, highly expressed proteins is observed. There are many more proteins in a given cell which are not visualized by such methods. Frequently it is the low abundance proteins that execute key regulatory functions." (page 1870, col. 1). Thus, Chen *et al.*, by selecting proteins detectable by staining of 2D gels, are likely to have excluded from their analysis many of the proteins most likely to be significant as cancer markers.

Secondly, Chen *et al.* looked at expression levels across a set of samples including a large number of tumor samples (76) along with a much smaller number of normal samples (9). The tumor samples were taken from stage 1 and stage III lung adenocarcinomas, which were classified as bronchoaveolar, bronchial derived or both bronchial and bronchoaveolar derived.

Accordingly, the tissues examined were from different tissues in different stages of normal or cancerous growth. The authors determined the relationship between mRNA and protein expression by using the average expression values for all samples. The average value for each protein or mRNA was generated using all 85 lung tissue samples. This resulted in negative normalized protein values in some cases. Further, the authors chose an arbitrary threshold of 0.115 for the correlation to be considered significant. Accordingly, the Chen paper does not account for different expression in different tissues or different stages of cancer.

Thirdly, no attempt was made to compare expression levels in normal versus tumor samples, and in fact the authors concede that they had too few normal samples for meaningful analysis (page 310, col. 2). As a result, the analysis in the Chen paper shows only that a number of randomly selected proteins have varying degrees of correlation between mRNA and protein expression levels within a set of different lung adenocarcinoma samples. The Chen paper does not address the issue of whether increased mRNA levels in the tumor samples taken together as one group, as compared to the normal samples as a group, correlated with increased protein levels in tumorous versus normal tissue. Accordingly, the results presented in the Chen paper are not applicable to the application at issue.

The correct test of utility is whether the utility is "more likely than not". In the case of the Chen reference, even if the analysis presented is correct (which is disputed), a review of the correlation coefficient data presented in the Chen *et al.* paper indicates that it is more likely than not that increased mRNA expression correlates with increased protein expression. A review of Table 1, which lists 66 genes [the paper incorrectly states there are 69 genes listed] for which only one protein isoform is expressed, shows that 40 genes out of 66 had a positive correlation between mRNA expression and protein expression. This clearly meets the test of "more likely than not." Similarly, in Table II , 30 genes with multiple isoforms [again the paper incorrectly states there are 29] were presented. In this case, for 22 genes out of 30, at least one isoform showed a positive correlation between mRNA expression and protein expression. Furthermore, 12 genes out of 29 showed a strong positive correlation [as determined by the authors] for at least one isoform. No genes showed a significant negative correlation. It is not surprising that not all isoforms are positively correlated with mRNA expression. Certain isoforms are likely

non-functional proteins. Thus, Table II also provides that it is more likely than not that protein levels will correlate with mRNA expression levels.

The same authors in Chen *et al.*, published a later paper, Beer *et al.*, Nature Medicine 8(8) 816-824 (2002) (copy enclosed as Exhibit A) which described gene expression of genes in adenocarcinomas and compared that to protein expression. In this paper they report that "these results suggest that the oligonucleotide microarrays provided reliable measures of gene expression" (page 817). The authors also state, "these studies indicate that many of the genes identified using gene expression profiles are likely relevant to lung adenocarcinoma." Clearly the authors of the Chen paper agree that microarrays provide a reliable measure of the expression levels of the gene and can be used to identify genes whose overexpression is associated with tumors.

Haynes et al. and Gygi et al.

The Examiner cites a new reference by Haynes *et al.* in support of the assertion that "mRNA (transcript) levels do not correlate with polypeptide levels in normal tissues." (Page 7 of the Examiner's Answer). Applicants respectfully point out that Haynes *et al.* never indicate that the correlation between mRNA and protein levels does not exist. Haynes *et al.* only state that "protein levels cannot be *accurately* predicted from the level of the corresponding mRNA transcript" (See page 1863, under Section 2.1, last line, emphasis added). This result is expected, since there are many factors that determine translation efficiency for a given transcript, or the half-life of the encoded protein. Not surprisingly, Haynes *et al.* concluded that protein levels cannot always be accurately predicted from the level of the corresponding mRNA transcript in a single cellular stage or type when looking at the level of transcripts across different genes.

Importantly, Haynes *et al.* did not say that for a single gene, a change in the level of mRNA transcript is not positively correlated with a change in the level of protein expression. Applicants have asserted that increasing the level of mRNA for a particular gene leads to a corresponding increase for the encoded protein. Haynes *et al.* did not study this issue and says absolutely nothing about it. One cannot look at the level of mRNA across several different genes to investigate whether a change in the level of mRNA for a particular gene leads to a change in the level of protein for that gene. Therefore, Haynes *et al.* is not inconsistent with or

contradictory to the utility of the instant claims, and offers no support for the PTO's rejection of Applicants' asserted utility.

Furthermore, Applicants note that contrary to the Examiner's statement, Haynes teaches that "*there was a general trend* but no strong correlation between protein [expression] and transcript levels" (See page 1863, under Section 2.1, emphasis added). For example, in Figure 1, there is a positive correlation between mRNA and protein amongst *most* of the 80 yeast proteins studied but the correlation is not linear, hence the authors suggest that one cannot *accurately* predict protein levels from mRNA levels. In fact, very few data points deviated or scattered away from the expected normal or showed a lack of correlation between mRNA: protein levels. Thus, the Haynes data meets the "more likely than not standard" and shows that a positive correlation exists between mRNA and protein. Therefore, Applicants submit that the Examiner's rejection is based on a misrepresentation of the scientific data presented in Haynes *et al.*

Haynes *et al.* may teach that protein levels cannot be "accurately predicted" from mRNA levels in the sense that the exact numerical amounts of protein present in a tissue cannot be determined based upon mRNA levels. Applicants respectfully submit that the PTO's emphasis on the need to "accurately predict" protein levels based on mRNA levels misses the point. The asserted utility for the claimed polypeptides is in the diagnosis of cancer. What is relevant to use as a cancer diagnostic is relative levels of gene or protein expression, not absolute values, that is, that the gene or protein is differentially expressed in tumors as compared to normal tissues. Applicants need only show that there is a correlation between mRNA and protein levels, such that mRNA overexpression generally predict protein overexpression. A showing that mRNA levels can be used to "accurately predict" the precise levels of protein expression is not required.

The Examiner also cites a new reference by Gygi *et al.*, a study on which the Haynes references is based. (Page 7 of the Examiner's Answer). Like Haynes, the Gygi reference looked at levels of mRNA at the same growth phase across different genes, not changes in mRNA levels for a single gene. Thus, when Gygi *et al.* state that "the correlation between mRNA and protein levels was insufficient to predict protein expression levels from quantitative mRNA data," the authors are referring to correlations between constant levels of mRNA and protein at the same growth phase across different genes, not a correlation between a change in mRNA level and a change in protein level for the same gene and corresponding protein.

Therefore, for the same reasons that Haynes is not relevant to Applicants' asserted utility, Gygi likewise offers no support for the PTO's rejection of Applicants' asserted utility.

Furthermore, Applicants submit that Gygi *et al* too did not indicate that a correlation between mRNA and protein levels does not exist. Gygi *et al.* only state that the correlation may not be sufficient in **accurately** predicting protein level from the level of the corresponding mRNA transcript (Emphasis added) (see page 1270, Abstract). *Accurate prediction* is not a criteria that is necessary for meeting the utility standards. Applicants note that the Gygi data indicate **a general trend** of correlation between protein [expression] and transcript levels (Emphasis added). For example, as shown in Figure 5, an mRNA abundance of **250-300** copies /cell correlates with a protein abundance of **500-1000** x 10^3 copies/cell. An mRNA abundance of **100-200** copies/cell correlates with a protein abundance of **250-500** x 10^3 copies/cell (emphasis added). Therefore, high levels of mRNA **generally** correlate with high levels of proteins. In fact, most data points in Figure 5 did not deviate or scatter away from the general trend of correlation. Thus, the Gygi data meets the "more likely than not standard" and shows that a positive correlation exists between mRNA and protein. Therefore, Applicants submit that the Examiner's rejection is based on a misrepresentation of the scientific data presented in Gygi *et al.*

Lian et al.

In further support of the alleged lack of correlation between mRNA expression and protein expression levels, the PTO has cited Lian *et al.* for the statement that there is a poor correlation between mRNA expression and protein abundance in mouse cells, and therefore it may be difficult to extrapolate directly from individual mRNA changes to corresponding ones in protein levels. (Page 8 of the Examiner's Answer).

In Lian *et al.*, the authors looked at the mRNA and protein levels of genes in a derived promyelocytic mouse cell-line during differentiation of the cells from a promyelocytic stage of development to mature neutrophils following treatment with retinoic acid. The level of mRNA expression was measured using 3'-end differential display (DD) and oligonucleotide chip array hybridization to examine the expression of genes at 0, 24, 48 and 72 hours after treatment with retinoic acid. Protein levels were qualitatively assessed at 0 and 72 hours after retinoic acid treatment following 2-dimensional gel electrophoresis.

Lian *et al.* report that they were able to identify 28 proteins which they considered differentially expressed (page 521). Of those 28, only 18 had corresponding gene expression information, and only 13 had measurable levels of mRNA expression (page 521, Table 6). The authors then compared the qualitative protein level from the 2-D electrophoresis gel to the corresponding mRNA level, and reported that only 4 genes of the 18 present in the database had expression levels which were consistent with protein levels (page 521, col. 1). The authors note that "[n]one of these was on the list of genes that were differentially expressed significantly (5-fold or greater change by array or 2-fold or greater change by DD)" (page 521; emphasis added). Based on these data, the authors conclude "[f]or protein levels based on estimated intensity of Coomassie dye staining in 2DE, there was poor correlation between changes in mRNA levels and estimated protein levels" (page 522, col. 2).

The authors themselves admit that there are a number of problems with the data presented in this reference. At page 520 of this article, the authors explicitly express their concerns by stating that "[t]hese data must be considered with several caveats: membrane and other hydrophobic proteins and very basic proteins are not well displayed by the standard 2DE approach, and proteins presented at low level will be missed. In addition, to simplify MS analysis, we used a Coomassie dye stain rather than silver to visualize proteins, and this decreased the sensitivity of detection of minor proteins." (emphasis added). It is known in the art that Coomassie dye stain is a very insensitive method of measuring protein. This suggests that the authors relied on a very insensitive measurement of the proteins studied. The conclusions based on such measurements can hardly be accurate or generally applicable. In particular, the total number of proteins examined by Lian *et al.* was only 50 (page 520, col. 2), as compared to the approximately 7000 genes for which mRNA levels were measured (page 515, col. 1). Thus the conclusions are based on a very small and atypical set of proteins.

Applicants also emphasize that Applicants are asserting that a measurable change in mRNA level generally leads to a corresponding change in the level of protein expression, not that changes in protein level can be used to predict changes in mRNA level. As discussed above, Lian *et al.* did not take genes which showed significant mRNA changes and check the corresponding protein levels. Instead, the authors looked at a small and unrepresentative number of proteins, and checked the corresponding mRNA levels. Based on the authors' criteria, mRNA

levels were significantly changed if they were at least 5-fold different when measured using a microchip array, or 2-fold different when using the more sensitive 3'-end differential display (DD). Of the 28 proteins listed in Table 6, only one has an mRNA level measured by microarray which is differentially expressed according to the authors (spot 7: melanoma X-actin, for which mRNA changed from 2539 to 341.3, and protein changed from 1 to 3). None of the other mRNAs listed in Table 6 show a significant change in expression level when using the criteria established by the authors for the less sensitive microarray technique.

There is also one gene in Table 6 whose expression was measured by the more sensitive technique of DD, and its level increased from a qualitative value of 0 to 2, a more than 2-fold increase (spot 2: actin, gamma, cytoplasmic). This increase in mRNA was accompanied by a corresponding increase in protein level, from 3 to 6.

Therefore, although the authors characterize the mRNA and protein levels as having a "poor correlation," this does not reflect a lack of a correlation between a change in mRNA level and a corresponding change in protein level. Only two genes meet the authors' criteria for differentially expressed mRNA level, and of those, one apparently shows a corresponding change in protein level and one does not. Thus, there is little basis for the authors' conclusion that "it may be difficult to extrapolate directly from individual mRNA changes to corresponding ones in protein levels (as estimated from 2DE)."

Finally, Applicants submit that Lian *et al.* only teach that protein expression may not correlate with mRNA level in differentiating myeloid cells and does not teach anything regarding such a lack of correlation for genes in general. Myeloid cell differentiation relates to hematopoiesis and is an entirely different biological process from solid tumor development because these two process involve entirely different regulatory mechanisms and molecules. Analysis of surface antigens expressed on myeloid cells of the granulocyte-monocyte-histiocyte series during differentiation in normal and malignant myelomonocytic cells is useful in identifying and classifying human leukemias and lymphomas, but cannot be used in diagnosis of any solid tumors. Therefore, even if the teaching of Lian *et al.* accurately reflects the correlation between mRNA and protein for the particular system studied, it can not apply to the tumor diagnosis assays of the present application.

Fessler et al.

The Examiner also cites a publication by Fessler *et al.*, as having "found a 'poor concordance between mRNA transcript and protein expression changes' in human cells." (Page 8 of the Examiner's Answer). Fessler is not contrary to Applicants' asserted utility, and actually supports Applicants' assertion that a change in the level of mRNA for a particular protein generally leads to a corresponding change in the level of the encoded protein. As noted above, Applicants make no assertions regarding changes in protein levels when mRNA levels are unchanged, nor does evidence of changes in protein levels when mRNA levels are unchanged have any relevance to Applicants' asserted utility.

Fessler *et al.* studied changes in neutrophil (PMN) gene transcription and protein expression following lipoplysaccharide (LPS) exposure. In Table VIII, Fessler *et al.* list a comparison of the change in the level of mRNA for 13 up-regulated proteins and 5 down-regulated proteins. Of the 13 up-regulated proteins, a change in mRNA levels is reported for only 3 such proteins. For these 3, mRNA levels are increased in 2 and decreased in the third. Of the 5 down-regulated proteins, a change in mRNA is reported for 3 such proteins. In all 3, mRNA levels also are decreased. Thus, in 5 of the 6 cases for which a change in mRNA levels are reported, the change in the level of mRNA corresponds to the change in the level of the protein. This is consistent with Applicants' assertion that a change in the level of mRNA for a particular protein generally leads to a corresponding change in the level of the encoded protein.

Regarding the remainder of the proteins listed in Table VIII, in 6 instances, protein levels changed while mRNA levels were unchanged. This evidence has no relevance to Applicants' assertion that changes in mRNA levels lead to corresponding changes in protein levels, since Applicants are not asserting that changes in mRNA levels are the only cause of changes in protein levels. In the final 6 instances listed in Table VIII, protein levels changed while mRNA was noted as "absent." This evidence also has no relevance to Applicants' assertion that changes in mRNA levels causes corresponding changes in protein levels. By virtue of being "absent," it is not possible to tell whether mRNA levels were increased, decreased or remained unchanged in PMN upon contact with LPS. Nothing in these results by Fessler *et al.* suggests that a change in the level of mRNA for a particular protein does not generally lead to a corresponding change in

the level of the encoded protein. Accordingly, these results are not contrary to Applicants' assertions.

The PTO points to Fessler's statement regarding Table VIII that there was "a poor concordance between mRNA transcript and protein expression changes." (Page 8 of the Examiner's Answer). As is clear from the above discussion, this statement does not relate to a lack of correlation between a change in mRNA levels leading to a change in protein levels, because in 5 of 6 such instances, changes in mRNA and protein levels correlated well. Instead, this statement relates to observations in which protein levels changed when mRNA was either unchanged or "absent." As such, this statement is an observation that in addition to transcriptional activity, LPS also has post-transcriptional and possibly post-translational activity that affect protein levels, an observation which is not contrary to Applicants' assertions. Accordingly, Fessler's results are consistent with Applicants' assertion that a change in mRNA level of for a particular protein generally leads to a corresponding change in the level of the encoded protein, since 5 of 6 genes demonstrated such a correlation.

Greenbaum et al.

In further support of the alleged lack of correlation between mRNA expression and protein expression levels, the Examiner cites an additional new reference by Greenbaum *et al.* The Examiner asserts that Greenbaum *et al.* teaches that, "To date, there have been only a handful of efforts to find correlations between mRNA and protein expression levels… And, for the most part, they have reported only minimal and/or limited correlations." (Page 8 of the Examiner's Answer).

Applicants note that Greenbaum *et al.* compared the expression of a number of different mRNAs and their corresponding proteins in yeast cells. Greenbaum *et al.* did not compare the change of expression of specific mRNAs and their corresponding proteins in cancer cells versus normal cells. Accordingly, this reference is also not relevant to the issue at hand. Nevertheless, Greenbaum states that logically "we would assume that those ORFs that show a large degree of variation in their expression are controlled at the transcriptional level. The variability of the mRNA expression is indicative of the cell controlling the mRNA expression at different points of the cell cycle to achieve the resulting and desired protein. **Thus we would expect and we found a high degree of correlation (r-0.89) between the reference mRNA and protein levels for**

these particular ORFs: the cell has already put significant energy into dictating the final level of protein through tightly controlling the mRNA expression" (page 117.5, col. 1; emphasis added). Furthermore, Greenbaum states that "we found that ORFs that have higher than average levels of ribosomal occupancy – that is that a large percentage of their cellular mRNA concentration is associated with ribosomes (being translated) – have well correlated mRNA and protein expression levels. (Figure 2)." (page 117.5, col. 2; emphasis added). Therefore, contrary to the Examiner's assertion, Greenbaum does find high levels of correlation between mRNA and protein expression in yeast cells. In particular, Greenbaum demonstrates that a high degree of correlation is found for those genes which show a large degree of variability in mRNA expression – that is, for those genes which show changes in mRNA expression, the change in mRNA expression is correlated with a change in protein expression.

In summary, Applicants respectfully submit that the Examiner has not shown that gene amplification in tumor as compared to normal tissue is not correlated with changes in mRNA and protein expression. The Patent Office has failed to meet its initial burden of proof that Applicants' claims of utility are not substantial or credible. The arguments presented by the Examiner in combination with the Pennica, Kònopka, Hu, Chen, Haynes, Gygi, Lian, Fessler, and Greenbaum articles do not provide sufficient reasons to doubt the statements by Applicants that PRO1293 has utility. As discussed above, the law does not require the existence of a "necessary" correlation between gene amplification and mRNA and protein expression levels. Nor does the law require that protein levels be "accurately predicted." According to the authors themselves, the data in the above cited references confirm that there is a general trend between gene amplification and mRNA and protein expression levels, which meets the "more likely than not standard" and show that a positive correlation exists between gene amplification and mRNA and protein expression. Therefore, Applicants submit that the Examiner's reasoning is based on a misrepresentation of the scientific data presented in the above cited reference and application of an improper, heightened legal standard. In fact, contrary to what the Examiner contends, the art indicates that, if a gene is overexpressed in cancer, it is more likely than not that the encoded protein will also be expressed at an elevated level.

It is "more likely than not" for amplified genes to have increased mRNA and protein levels

Applicants have submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is more likely than not that the encoded protein will be expressed at an elevated level. First, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.*, (made of record in Applicants' Response filed August 19, 2004) collectively teach that in general, gene amplification increases mRNA expression. Second, the Declaration of Dr. Paul Polakis, principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, shows that, in general, there is a correlation between mRNA levels and polypeptide levels.

The Examiner has asserted that "Orntoft et al. could only compare the levels of about 40 well-resolved and focused abundant proteins." (Page 18 of the Examiner's Answer). Applicants respectfully point out that while technical considerations did prevent Orntoft *et al.* from evaluating a larger number of proteins, the ones they did look at showed a clear correlation between mRNA and protein expression levels. The authors found that "[i]n general **there was a highly significant correlation ($p<0.005$) between mRNA and protein alterations**. Only one gene [of the 40 examined] showed disagreement between transcript alteration and protein alteration" (page 42, col. 2; emphasis added). Clearly, a correlation in 39 of 40 genes examined supports Applicants' assertion that changes in mRNA level generally lead to corresponding changes in protein level.

The Examiner further asserts that "Applicants have provided no fact or evidence concerning a lack of correlation between the specification's disclosure of low levels of amplification of DNA (which were not characterized on the basis of those in the Orntoft publication) and an associated rise in level of the encoded protein." (Page 18 of the Examiner's Answer).

As discussed above, the levels of amplification for PRO1293 were **not** "low" but significant, and ranged from 2.19-fold to 5.03-fold, in three different lung and colon tumors. Applicants note that the levels of gene amplification observed by Orntoft *et al.* were relatively low, averaging only 0.3-0.4-fold (page 40, col. 1). In particular, the level of gene amplification associated with expression changes was only around two-fold (see Figure 2), even less than the

2.19-fold to 5.03-fold amplification observed for PRO1293. Even with these relatively low levels of gene amplification, Orntoft *et al.* found that "**[i]n most cases**, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%)" (page 40, col. 2; emphasis added). The level of correlation between DNA copy number and increased mRNA levels observed by Orntoft *et al.*, from **77-80%**, clearly meets the standard of more likely than not. Orntoft *et al.* also found a "highly significant" correlation between mRNA and protein levels, with the two data sets studied having correlations of 39/40 (**98%**) and 19/26 (**73%**) (pages 42-43).

The Examiner also states that Orntoft *et al.* do not compare gene expression in cancerous versus non-cancerous tissue, and thus "Orntoft *et al.* did not find any cancer markers." (Page 21 of the Examiner's Answer). Applicants note that while Orntoft *et al.* did not compare cancerous versus non-cancerous tissues, they did compare invasive versus benign tumors, thus finding genes that were markers of tumor malignancy.

Applicants respectfully submit that the Examiner also appears to misunderstand the data presented by Hyman *et al.* The Examiner asserts that "of the 12,000 transcripts analyzed, a set of 270 was identified in which overexpression was attributable to gene amplification." The Examiner concludes that "[t]his proportion is 2%; the Examiner maintains that 2% does not provide a reasonable expectation that the slight amplification of PRO1293 would be correlated with elevated levels of mRNA." (Page 18 of the Examiner's Answer). Appellants respectfully submit that the Examiner appears to have misinterpreted the results of Hyman *et al.* Hyman *et al.* chose to do a genome-wide analysis of a large number of genes, most of which, as shown in Figure 2, were not amplified. Accordingly, the 2% number is meaningless, as the low figure mainly results from the fact that only a small percentage of genes are amplified in the first place. The significant figure is not the percentage of genes in the genome that show amplification, but the percentage of amplified genes that demonstrate increased mRNA and protein expression.

The Examiner further asserts that the Hyman reference "found 44% of *highly* amplified genes showing overexpression at the mRNA level, and 10.5% of highly overexpressed genes being amplified; thus, even at the level of high amplification and high overexpression, the two do not correlate." (Page 18 of the Examiner's Answer). Applicants submit that the 10.5% figure is not relevant to the issue at hand. One of skill in the art would understand that there can be more

than one cause of overexpression. The issue is not whether overexpression is always, or even typically caused by gene amplification, but rather, whether gene amplification typically leads to overexpression.

The Examiner's assertion is not consistent with the interpretation Hyman *et al.* themselves place on their data, stating that, "The results illustrate **a considerable influence of copy number on gene expression patterns**." (page 6242. col. 1; emphasis added). In the more detailed discussion of their results, Hyman *et al.* teach that "[u]p to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (***i.e.*, belonged to the global upper 7% of expression ratios**) compared with only 6% for genes with normal copy number." (See page 6242, col. 1; emphasis added). These details make it clear that Hyman *et al.* set a highly restrictive standard for considering a gene to be overexpressed; yet almost half of all highly amplified transcripts met even this highly restrictive standard. Therefore, the analysis performed by Hyman *et al.* clearly shows that it is "more likely than not" that a gene which is amplified in tumor cells will have increased gene expression.

The Examiner asserts that Hyman *et al.* and Pollack *et al.* do not examine protein expression. (Page 18 of the Examiner's Answer). Applicants submit that the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* were submitted primarily as evidence that in general, gene amplification increases mRNA expression. As evidence that, in general, there is a correlation between mRNA levels and polypeptide levels, Applicants further submitted the Declaration of Dr. Paul Polakis. Thus Applicants do not rely upon the Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* articles to show a correlation between mRNA levels and polypeptide levels, because such a correlation is demonstrated in the Polakis Declaration. Nonetheless, as discussed above, Orntoft *et al.* does provide evidence that increased mRNA levels in tumor cells are associated with increased protein levels in the same tumor cells.

Finally, the Examiner asserts that "Pollack *et al.* is similarly limited to highly amplified genes which were not evaluated by the method of the instant specification." The Examiner further notes that none of the three references is directed to lung or colon cancer. (Page 18 of the Examiner's Answer). Applicants note that, as discussed above, the levels of amplification for PRO1293 were **not** "low" but **significant**. Applicants further respectfully submit that the Examiner has provided no arguments or evidence as to why the data from Orntoft *et al.*, Hyman

et al. and Pollack *et al.*, concerning gene expression in bladder and breast tumors, would not also apply to tumors in general.

With regard to the correlation between mRNA expression and protein levels, the Examiner has asserted that the Polakis Declaration is insufficient to overcome the rejection of the claims since it is limited to a discussion of data regarding the correlation of mRNA levels and polypeptide levels and not gene amplification levels. The Examiner further asserts that there is "strong opposing evidence showing that gene amplification is not predictive of increased mRNA levels in normal tissues and, in turn, that increased mRNA levels are frequently not predictive of increased polypeptide levels." (Pages 19-20 of the Examiner's Answer).

Applicants submit that Dr. Polakis' Declaration was presented to support the position that there is a correlation between mRNA levels and polypeptide levels, the correlation between gene amplification and mRNA levels having already been established by the data shown in the Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* articles. With regard to the alleged "strong opposing evidence" that increased mRNA levels are not predictive of increased polypeptide levels, Applicants have discussed in detail above the reasons why the data in the Hu, Chen, Haynes, Gygi, Lian, Fessler, and Greenbaum articles confirm that there is a general trend between mRNA and protein expression levels, which meets the "more likely than not standard" and show that a positive correlation exists between mRNA expression and protein expression.

The Examiner asserts that "the data are not included in the declaration so that the examiner could not independently evaluate them." (Page 20 of the Examiner's Answer). Applicants emphasize that the opinions expressed in the Polakis Declaration are all based on factual findings. Thus, Dr. Polakis explains that in the course of their research using microarray analysis, he and his co-workers identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. Subsequently, antibodies binding to about 30 of these tumor antigens were prepared, and mRNA and protein levels were compared. In approximately 80% of the cases, the researchers found that increases in the level of a particular mRNA correlated with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells. Dr. Polakis' statement that "an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the

tumor cell relative to the normal cell" is based on factual, experimental findings, clearly set forth in the Declaration. Accordingly, the Declaration is not merely conclusive, and the fact-based conclusions of Dr. Polakis would be considered reasonable and accurate by one skilled in the art.

Furthermore, without acquiescing to the propriety of this rejection, and merely to expedite prosecution in this case, **Applicants present a second Declaration by Dr. Polakis (Polakis II) that presents evidentiary data in Exhibit B**. Exhibit B of the Declaration identifies 28 gene transcripts out of 31 gene transcripts (i.e., greater than 90%) that showed good correlation between tumor mRNA and tumor protein levels. As Dr. Polakis' Declaration (Polakis II) says "[a]s such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA." Accordingly, Dr. Polakis has provided the facts to enable the Examiner to draw independent conclusions.

The case law has clearly established that in considering affidavit evidence, the Examiner must consider all of the evidence of record anew.[1] "After evidence or argument is submitted by the applicant in response, patentability is determined on the totality of the record, by a preponderance of the evidence with due consideration to persuasiveness of argument."[2] Furthermore, the Federal Court of Appeals held in *In re Alton*, "We are aware of no reason why opinion evidence relating to a fact issue should not be considered by an Examiner."[3] Applicants also respectfully draw the Examiner's attention to the Utility Examination Guidelines[4] which state, "Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered." The statement in question from an expert in the field (the Polakis Declaration) states: "it is my

---

[1] *In re Rinehart,* 531 F.2d 1084, 189 U.S.P.Q. 143 (C.C.P.A. 1976); *In re Piasecki,* 745 F2d. 1015, 226 U.S.P.Q. 881 (Fed. Cir. 1985).

[2] *In re Alton,* 37 U.S.P.Q.2d 1578, 1584 (Fed. Cir 1996) (quoting *In re Oetiker,* 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992)).

[3] *Id.* at 1583.

[4] Part IIB, 66 Fed. Reg. 1098 (2001).

considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell." Therefore, barring evidence to the contrary regarding the above statement in the Polakis declaration, this rejection is improper under both the case law and the Utility guidelines.

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is a correlation between polypeptide and mRNA levels, these instances are exceptions rather than the rule. In the majority of amplified genes, the teachings in the art, as exemplified by Orntoft *et al.,* Hyman *et al.,* Pollack *et al.,* and the Polakis Declarations, overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO1293 gene, that the PRO1293 polypeptide is concomitantly overexpressed. Thus, Applicants submit that the PRO1293 polypeptide, and the claimed antibodies that bind it, have utility in the diagnosis of cancer and based on such a utility, one of skill in the art would know exactly how to use the claimed antibodies for diagnosis of cancer.

Accordingly, Applicants request the Examiner to reconsider and withdraw the rejection of Claims 28-32 under 35 U.S.C. §§101 and 112.

CONCLUSION

In conclusion, the present application is believed to be in *prima facie* condition for allowance, and an early action to that effect is respectfully solicited. Should there be any further issues outstanding, the Examiner is invited to contact the undersigned agent at the telephone number shown below.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. **08-1641** (referencing Attorney's Docket No. **39780-2830 P1C4**).

Respectfully submitted,

Date: June2, 2006

By: _____
Barrie D. Greene (Reg. No. 46,740)

HELLER EHRMAN LLP
275 Middlefield Road
Menlo Park, California 94025
Telephone: (650) 324-7000
Facsimile: (650) 324-0638

SV 2211207 v1
6/2/06 11:50 AM (39780.2830)

AF
ZJW

Please type a plus sign (+) inside this box ☐ [+]

06-05-06

PTO/SB/21 (6-99)
Approved for use through 09/30/2000. OMB 0651-0031
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	10/006,818
Filing Date	December 6, 2001
First Named Inventor	Kevin P. Baker
Group/Art Unit	1647
Examiner Name	Fozia M. Hamud
Attorney Docket Number	39780-2830 P1C4

Total Number of Pages in This Submission	55

ENCLOSURES *(check all that apply)*

- ☒ **Fee Transmittal Form**
  - ☐ Fee Attached
- ☐ Amendment/Response
  - ☐ After Final
  - ☐ Version With Markings Showing Changes
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☒ **Information Disclosure Statement with 1 Reference**
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Copy of Notice
- ☐ Copy of an Assignment
- ☐ Drawing(s)
- ☒ **Declaration under 37 CFR §1.132 for Paul Polakis**
- ☐ Petition Routing Slip (PTO/SB/69) and Accompanying Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, by Assignee to Exclusion of Inventor Under 37 C.F.R. §3.71 With Revocation of Prior Powers
- ☐ Terminal Disclaimer
- ☐ Small Entity Statement
- ☐ Request for Refund
- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☒ **Appeal Communication to Group** ***(Appeal Notice, Brief, Reply Brief, Request for Reopening of Prosecution and Response under 37 CFR §1.111))***
- ☐ Proprietary Information
- ☐ Status Letter
- ☒ **ADDITIONAL ENCLOSURE(S)** ***(PLEASE IDENTIFY BELOW):***
  - ☒ **STAMPED RETURN POSTCARD**

Remarks

AUTHORIZATION TO CHARGE DEPOSIT ACCOUNT 08-1641 FOR ANY FEES DUE IN CONNECTION WITH THIS PAPER, REFERENCING ATTORNEY'S DOCKET NO. 39780-2830P1C4.

SIGNATURE OF APPLICANT, ATTORNEY OR AGENT

Firm *or* Individual name	**HELLER EHRMAN LLP**	**BARRIE D. GREENE (Reg. No. 46,740)**	
	275 Middlefield Road, Menlo Park, California 94025	**Telephone: (650) 324-7000**	**Facsimile: (650) 324-0638**
Signature			
Date	**JUNE 2, 2006**	Customer Number:	**35489**

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. §1.10 on the date indicated below and addressed to: **MAIL STOP APPEAL BRIEF -PATENTS,** Commissioner for Patents, PO Box 1450, Alexandria, Virginia 22313-1450, on this date: **JUNE 2, 2006**

Express Mail Label **EV 765 973 322 US**

Typed or printed name	**L. ACOSTA**		
Signature		Date	**JUNE 2, 2006**

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop ___, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

SECOND DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I am currently employed by Genentech, Inc. where my job title is Staff Scientist.

2. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As I stated in my previous Declaration dated May 7, 2004 (attached as Exhibit A), my laboratory has been employing a variety of techniques, including microarray analysis, to identify genes which are differentially expressed in human tumor tissue relative to normal human tissue. The primary purpose of this research is to identify proteins that are abundantly expressed on certain human tumor tissue(s) and that are either (i) not expressed, or (ii) expressed at detectably lower levels, on normal tissue(s).

4. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor tissue at significantly higher levels than in normal human tissue. To date, we have successfully generated antibodies that bind to 31 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human tumor tissue and normal tissue. We have then quantitatively compared the levels of mRNA and protein in both the tumor and normal tissues analyzed. The results of these analyses are attached herewith as Exhibit B. In Exhibit B, "+" means that the mRNA or protein was detectably overexpressed in the tumor tissue relative to normal tissue and "-" means that no detectable overexpression was observed in the tumor tissue relative to normal tissue.

5. As shown in Exhibit B, of the 31 genes identified as being detectably overexpressed in human tumor tissue as compared to normal human tissue at the mRNA level, 28 of them (i.e., greater than 90%) are also detectably overexpressed in human tumor tissue as compared to normal human tissue at the protein level. As such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4-5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor tissue relative to a normal tissue more often than not correlates to a similar increase in abundance of the encoded protein in the tumor tissue relative to the normal tissue. In fact, it remains a generally accepted working assumption in molecular biology that increased mRNA levels are more often than not predictive of elevated levels of the encoded protein. In fact, an entire industry focusing on the research and development of therapeutic antibodies to treat a variety of human diseases, such as cancer, operates on this working assumption.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 3-29-06

By: Paul Polakis

Paul Polakis, Ph.D.

EXHIBIT A

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).

2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.

4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.

5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

SV 2031808 v1

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South SanFrancisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985	Assistant Professor, Department of Chemistry, Oberlin College, Oberlin, Ohio
1980-1984	Graduate Research Assistant, Department of Biochemistry, Michigan State University East Lansing, Michigan

PUBLICATIONS:

1. **Polakis, P G.** and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.

2. **Polakis, P.G.** and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.

3. **Polakis, P. G.** and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.

4. Uhing, R.J., **Polakis,P.G.** and Snyderman, R. 1987 Isolaton of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.

5. **Polakis, P.G.**, Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.

6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. **Polakis, P.G.**, Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.

8. **Polakis, P. G.**, Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Comun.** 160, 25-32.

9. **Polakis, P.**, Weber,R.F., Nevins,B., Didsbury, J. Evans,T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.G.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Charaterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K^+ Channels. **Cell.** 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.

14. Polakis, P.G. Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits $p21^{ras}$ Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein $p21^{rap1}$. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for $p21^{ras}$-dependent Inhibition of Muscarinic Atrial K^+ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P** . 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 $p60^{v\text{-}src}$ Can Associate with and Phosphorylate the $p21^{ras}$ GTPase Activating Protein. **Oncogene** 7,389-394.

22. Polakis P. and McCormick, F. 1992 Interactions Between $p21^{ras}$ Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., **Polakis,P.** and McCormick, F. 1992 Molecular coloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.

24. **Polakis, P.**, Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle $p34^{cdc2}$ Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., **Polakis, P.**, McCormick, F., and Innis, M. A. 1992 Functional Interactions Between $p21^{rap1A}$ and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy,L., Clark, R., McCormick, F. and **Polakis, P.** 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell . Biol.** 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., **Polakis, P.**, McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.

28. Janoueix-Lerosey, I., **Polakis, P.**, Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.

29. **Polakis, P.** 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the *ras*-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. **Polakis, P.** and McCormick, F. 1993 Structural requirements for the interaction of $p21^{ras}$ with GAP, exchange factors, and its bological effector target. **J. Biol Chem.** 268, 9157-9160.

31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and **Polakis, P.** 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., **Polakis, P.**, McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.

33. Sato, K. Y., **Polakis, P.**, Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor acitvity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., **Polakis, P.** and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and **Polakis, P.** 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP.. **Methods Enz.** 255,31

37. **Polakis, P.** 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71

38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α-catenin, β-catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555

39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β-catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.

40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.

41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.

42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.

43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β-catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.**16, 4088-4094.

44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein witha rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.

45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.

46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.

47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635

48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in Xenopus embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β-catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. **Polakis, P.** The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P** 1997 Loss of β-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes. K., Waterman, M., and **Polakis, P.** 1997 Induction of a β-catenin-LEF-1 complex by wnt-1 and transforming mutants of β-catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β-catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β-catenin and glycogen synthase kinase 3β. **Current Biology 8,** 573-581.

56. **Polakis, P.** 1998 The oncogenic activation of β-catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfled, Florence Margottin, Richard Benarous and **Paul Polakis.** 1999 The F-box protein β–TrCP associates with phosphorylated β–catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β–catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P**, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β-catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.

61 **Polakis P**, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin
in colorectal cancer. **Adv Exp Med Biol**. 470, 23-32

62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene**. 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett**. 458, 247-51.

65. Peifer M, **Polakis P**. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.

66. **Polakis P**. 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.

67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.

68. Szeto , W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A.,
Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-
Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and
Retinoic Acid. **Cancer Res** 61, 4197-4204.

69. Rubinfeld B, Tice DA, **Polakis P**. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem**
276, 39037-39045.

70. **Polakis P**. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.

71. Tice DA, Soloviev I, **Polakis P**. 2002 Activation of the Wnt Pathway Interferes withSerum Response Element-driven Transcription of Immediate Early Genes. **J Biol.**
Chem. 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennnica D, **Polakis P.** 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.

73. **Polakis, P.** 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao,W. , Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., **Polakis, P.** 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, **Polakis P,** Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics,** April 8

EXHIBIT B

	tumor mRNA	tumor IHC
UNQ2525	+	+
UNQ2378	+	+
UNQ972	+	-
UNQ97671	+	+
UNQ2964	+	+
UNQ323	+	+
UNQ1655	+	+
UNQ2333	+	+
UNQ9638	+	+
UNQ8209	+	+
UNQ6507	+	+
UNQ8196	+	+
UNQ9109	+	+
UNQ100	+	+
UNQ178	+	+
UNQ1477	+	+
UNQ1839	+	+
UNQ2079	+	+
UNQ8782	+	+
UNQ9646	+	-
UNQ111	+	+
UNQ3079	+	+
UNQ8175	+	+
UNQ9509	+	+
UNQ10978	+	-
UNQ2103	+	+
UNQ1563	+	+
UNQ16188	+	+
UNQ13589	+	+
UNQ1078	+	+
UNQ879	+	+

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	)	Examiner: Hamud, Fozia M.
Kevin P. BAKER, et al.	)	Art Unit: 1647
Application Serial No. 10/006,818	)	Confirmation No: 1321
Filed: December 6, 2001	)	Attorney's Docket No. 39780-2830 P1C4
For: **SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME**	)	**Customer No. 35489**

EXPRESS MAIL LABEL NO. EV 765 973 322 US
DATE MAILED: JUNE 2, 2006

INFORMATION DISCLOSURE STATEMENT UNDER 37 C.F.R. §1.97

MAIL STOP RCE
Commissioner for Patents
PO Box 1450
Alexandria, Virginia 22313-1450

Sir:

Listed below or on an attached Form PTO-1449 is information known to applicant(s). A copy of each listed publication and U.S. and foreign patent, except for pending U.S. applications, is being submitted herewith, along with a concise explanation of information in a foreign language, if any, pursuant to 37 C.F.R. §1.97-1.98.

Applicants respectfully request that the listed information be considered by the Examiner and be made of record in the above-identified application. If Form PTO-1449 is enclosed, the Examiner is requested to initial and return it in accordance with M.P.E.P. §609.

This statement is not intended to represent that a search has been made or that the information cited in the statement is, or is considered to be, material to patentability as defined in §1.56.

☐ This statement qualifies under *37 C.F.R. §1.97*, subsection (b) because (check all that apply):

☐ (1) It is being filed within 3 months of the application filing date and is other than a continued prosecution application under § 1.53(d)

-- OR --

☐ (2) It is being filed within 3 months of entry of a national stage

-- OR --

☐ (3) It is being filed before the mail date of the first Office Action on the merits

-- OR --

☐ (4) It is being filed before the mailing of a first Office Action after the filing of a request for continued examination under § 1.114.

☒ *37 C.F.R. §1.97(c).* If this statement is being filed after the latest of: (1) three months beyond the filing date of a national application; (2) three months beyond the date of entry of the national stage as set forth in §1.491 in an international application; or (3) the mailing date of a first Office action on the merits, but before the mailing date of the earlier of a final office action under §1.113 or a notice of allowance under §1.311, then:

☐ a certification as specified in §1.97(e) is provided below; **or**

☒ a fee of **$180.00** as set forth in §1.17(p) is authorized below, enclosed, or included with the payment of other papers filed together with this statement.

☐ *37 C.F.R. §1.97(d).* If this statement is being filed after the mailing date of the earlier of a final office action under §1.113 or a notice of allowance under §1.311, but before payment of the issue fee, then:

A. a certification as specified in §1.97(e) is completed below; and

B. a petition under 37 C.F.R. §1.97(d) requesting consideration of this statement is submitted herewith; **and**

C. a fee of $130.00 as set forth in §1.17(i)(1) is authorized below, enclosed, or included with the payment of other papers filed together with this statement.

☒ *Fee Authorization.* The Commissioner is hereby authorized to charge the above-referenced fees of **$180.00** and charge any additional fees or credit any overpayment associated with this communication to Deposit Account No. **08-1641 (Attorney's Docket No. 39780-2830 P1C4)**

Respectfully submitted,

Dated: June 2, 2006

By: _____
Barrie D. Greene (Reg. No. 46,740)

HELLER EHRMAN LLP
275 Middlefield Road
Menlo Park, California 94025-3506
Telephone: (650) 324-7000
Facsimile: (650) 324-0638

Form PTO-1449	Attorney's Docket No.	Application Serial No.
INFORMATION DISCLOSURE STATEMENT	39780-2830 P1C4	10/006,818
	Applicant(s)	Kevin P. BAKER, et al.
(use several sheets if necessary)	Filing Date: December 6, 2001	Group Art Unit: 1647

U.S. PATENT DOCUMENTS

Examiner Initials	Ref. No.	Date	Document No.	Name	Class	Subclass	Filing Date (if appropriate)

FOREIGN PATENT DOCUMENTS

Examiner Initials	Ref. No.	Date	Document No.	Name	Class	Subclass	Translation YES	Translation NO

OTHER DOCUMENTS
(including author, title, date, pertinent pages, etc.)

Examiner Initials	Ref. No.	Title
	1.	Beer, D.G., et al., "Gene-Expression Profiles Predict Survival of Patients With Lung Adenocarcinoma," *Nature Medicine* - 8(8):816-824 (2002)

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if citation considered, whether or not the citation conforms with MPEP 609. Draw a line through the citation if not in conformance and not considered. Include a copy of this form with next communication to applicant.

*If an asterisk is placed beside the reference number, a copy is not provided because the reference was previously cited by or submitted to the PTO in a prior application that is identical in the statement and relied upon for an earlier filing date under 35 U.S.C. §120. 37 C.F.R. §1.98 (d).

Gene-expression profiles predict survival of patients with lung adenocarcinoma

DAVID G. BEER[1], SHARON L.R. KARDIA[2], CHIANG-CHING HUANG[3], THOMAS J. GIORDANO[4], ALBERT M. LEVIN[2], DAVID E. MISEK[5], LIN LIN[1], GUOAN CHEN[1], TAREK G. GHARIB[1], DAFYDD G. THOMAS[4], MICHELLE L. LIZYNESS[4], RORK KUICK[5], SATORU HAYASAKA[3], JEREMY M.G. TAYLOR[3], MARK D. IANNETTONI[1], MARK B. ORRINGER[1] & SAMIR HANASH[5]

Departments of [1]Surgery, [2]Epidemiology, [3]Biostatistics, [4]Pathology and [5]Pediatrics, University of Michigan, Ann Arbor, Michigan, USA
Correspondence should be addressed to D.G.B.; email: dgbeer@umich.edu.

Published online: 15 July 2002, doi:10.1038/nm733

Histopathology is insufficient to predict disease progression and clinical outcome in lung adenocarcinoma. Here we show that gene-expression profiles based on microarray analysis can be used to predict patient survival in early-stage lung adenocarcinomas. Genes most related to survival were identified with univariate Cox analysis. Using either two equivalent but independent training and testing sets, or 'leave-one-out' cross-validation analysis with all tumors, a risk index based on the top 50 genes identified low-risk and high-risk stage I lung adenocarcinomas, which differed significantly with respect to survival. This risk index was then validated using an independent sample of lung adenocarcinomas that predicted high- and low-risk groups. This index included genes not previously associated with survival. The identification of a set of genes that predict survival in early-stage lung adenocarcinoma allows delineation of a high-risk group that may benefit from adjuvant therapy.

Lung cancer remains the leading cause of cancer death in industrialized countries. Most patients with non-small cell lung cancer (NSCLC) present with advanced disease, and despite recent advances in multi-modality therapy, the overall 10-year survival rate remains a dismal 8–10%[1]. However, a significant minority of patients (~25–30%) with NSCLC have stage I disease and receive surgical intervention alone. Although 35–50% of patients with stage I disease will relapse within 5 years[2–4], it is not currently possible to identify specific high-risk patients.

Adenocarcinoma is currently the predominant histological subtype of NSCLC (refs. 1,5,6). Although morphological assessment of lung carcinomas can roughly stratify patients, there is a need to identify patients at high risk for recurrent or metastatic disease. Preoperative variables that affect survival of patients with NSCLC have been identified[7–10]. Tumor size, vascular invasion, poor differentiation, high tumor-proliferative index and several genetic alterations, including K-*ras* (refs. 11,12) and p53 (refs. 10,13) mutations, have prognostic significance. Multiple independently assessed genes or gene products have also been investigated to better predict patient prognosis in lung cancer[14–18]. Technologies that simultaneously analyze the expression of thousands of genes[19] can be used to correlate gene-expression patterns with numerous clinical parameters—including patient outcome—to better predict tumor behavior in individual patients[20]. Analyses of lung cancers using array technologies have identified subgroups of tumors that differ according to tumor type and histological subclasses and, to a lesser extent, survival among adenocarcinoma patients[21,22]. Here we correlated gene-expression profiles with clinical outcome in a cohort of patients with lung adenocarcinoma and identified specific genes that predict survival among patients with stage I disease. For further validation, we also show that the risk index predicted survival in an independent cohort of stage I lung adenocarcinomas.

Hierarchical profile clustering yields three tumor subsets

Using oligonucleotide arrays, we generated gene-expression profiles for 86 primary lung adenocarcinomas, including 67 stage I and 19 stage III tumors, as well as 10 non-neoplastic lung samples. Selected sample replicates showed high correlation among coefficients and reliable reproducibility. We determined transcript abundance using a custom algorithm and the data set was trimmed of genes expressed at extremely low levels, that is, genes were excluded if the measure of their 75th percentile value was less than 100. Although potentially resulting in the loss of some information, trimming in this manner decreased the possibility that the clustering algorithm would be strongly influenced by genes with little or no expression in these samples. Hierarchical clustering with the resulting 4,966 genes yielded 3 clusters of tumors (Fig. 1). All 10 non-neoplastic samples clustered tightly together within Cluster 1 (data not shown). We examined the relationships between cluster and patient and tumor characteristics (Fig. 1 and Supplementary Figure A online). There were associations between cluster and stage ($P = 0.030$) and between cluster and differentiation ($P = 0.01$). Cluster 1 contained the greatest percentage (42.8%) of well differentiated tumors, followed by Cluster 2 (27%) and Cluster 3 (4.7%). Cluster 3 contained the highest percentage of both poorly differentiated (47.6%) and stage III tumors (42.8%), yet contained 3 (14.3%) moderately differentiated and 1 (5%) well differentiated stage I tumor. Notably, 11 stage I tumors were present in Cluster 3, sug-

gesting a common gene-expression profile for this subset of stage I and stage III tumors.

For patients with stage I and stage III tumors, the average ages were 68.1 and 64.5 years and the percentage of smokers was 88.9% and 89.5%, respectively. Marginally significant associations between cluster and smoking history were observed ($P = 0.06$). A significant relationship between histopathological classification and cluster was only discernable for bronchioloalveolar adenocarcinomas (BAs), which were only present in Clusters 1 and 2 ($P = 0.0055$) and comprised 35.7% and 12.3% of tumors for Clusters 1 and 2, respectively.

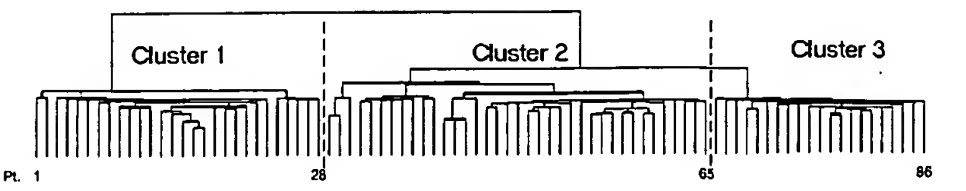


Fig. 1 Unsupervised classification analysis of lung adenocarcinomas. 3 classes of tumors identified by agglomerative hierarchical clustering of gene-expression profiles using the 4,966 expressed genes. Patient and histopathological information for each lung adenocarcinoma case by cluster designation and methods for K-*ras* 12/13th-codon mutational status and nuclear p53 protein accumulation are provided (Supplementary Figure A online). TN classification denotes information regarding patient tumor size and nodal involvement. Associations between cluster membership and patient or histopathological variables are indicated at significance level ($P \leq 0.05$).

We examined the heterogeneity in gene-expression profiles based on the trimmed data set among normal lung samples and stage I and stage III adenocarcinomas by calculating correlation coefficients between all pairs of samples. In contrast to normal lung samples that displayed highly similar gene-expression profiles (median correlation, 0.9), both stage I and III lung tumors demonstrated much greater heterogeneity in their expression profiles with lower correlation coefficients (median values, 0.82 and 0.79, respectively).

Northern-blot and immunohistochemistry analyses

Of the 4,966 genes examined, 967 differed significantly between stage I and III adenocarcinomas, a number in excess of that expected by chance alone (248 at alpha level (α) = 0.05). Three genes were arbitrarily selected to verify the microarray expression data. The mRNA from 20 of the normal lung and tumor samples was examined by northern-blot hybridization with probes for insulin-like growth factor–binding protein 3 (*IGFBP3*), cystatin C and lactate dehydrogenase A (*LDH-A*) (Fig. 2*a*). Two gene probes not represented on the microarrays were used as controls, including histone H4, a potential index of overall cell proliferation, and 28S ribosomal RNA, a control for sample loading and transfer. The relative amounts of *IGFBP3*, cystatin C and *LDH-A* mRNA strongly correlated with microarray-based measurements (Fig. 2*b*). In both assays, *IGFBP3* and *LDH-A* mRNA levels increased from stage I to stage III adenocarcinomas and were higher than those in normal lung. Cystatin C mRNA levels were more variable but relatively greater in normal lung than tumors. These results suggest that the oligonucleotide microarrays provided reliable measures of gene expression. The tumors showed slightly greater histone H4 expression than the normal lung, likely reflecting increased proliferation of tumor cells.

Immunohistochemistry was performed for IGFBP3, cystatin C and HSP-70 to determine whether mRNA overexpression was reflected by an increase of their corresponding proteins in tumors.

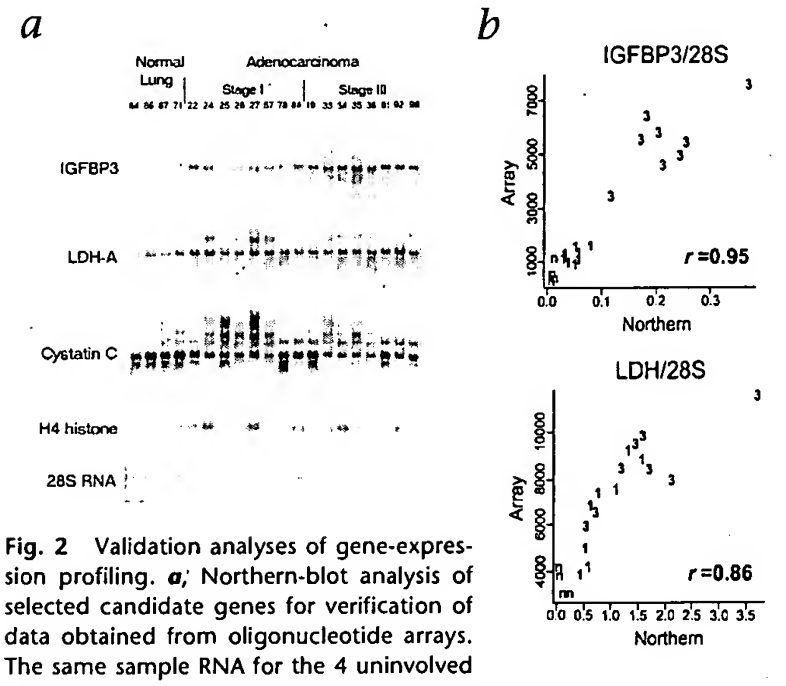
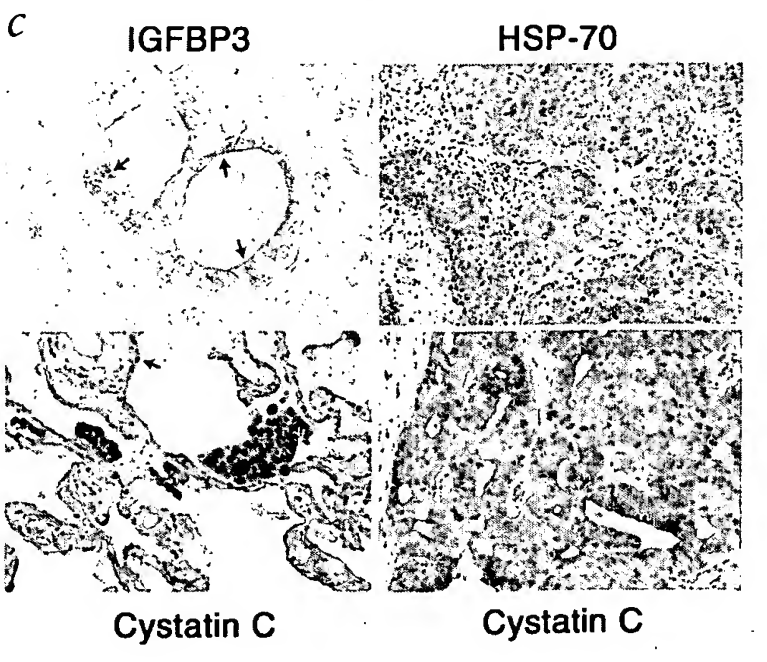


Fig. 2 Validation analyses of gene-expression profiling. ***a***, Northern-blot analysis of selected candidate genes for verification of data obtained from oligonucleotide arrays. The same sample RNA for the 4 uninvolved lung, 8 stage I and 8 stage III tumors was used for the northern-blot and oligonucleotide array analyses. ***b***, Correlation analysis of quantitative data obtained from oligonucleotide arrays and northern blots measured by integrated phosphorimager-based signals for the *IGFBP3* and *LDH-A* genes. The ratio of *IGFBP3*, cystatin C and *LDH-A* mRNA to *28S* rRNA was determined. The relative values for each gene from each sample are shown. n, non-neoplastic normal lung; 1, stage I tumors; 3, stage III tumors. ***c***, Immunohistochemical analysis of IGFBP-3, HSP-70 and cystatin C in lung and lung adenocarcinomas. Cytoplasmic IGFBP-3 immunoreactivity in a neoplastic gland (tumor L22) with prominent apical staining (blue reactant staining, arrow, upper left). Diffuse cytoplasmic HSP-70 immunoreactivity (tumor L27), yet stromal elements show no reactivity (upper right). Normal lung parenchyma (lower left) shows cytoplasmic cystatin C immunoreactivity in alveolar pneumocytes (arrow) and intra-alveolar macrophages but tumor (L90) shows diffuse cytoplasmic cystatin C immunoreactivity with prominent apical staining (lower right). Magnification, ×200



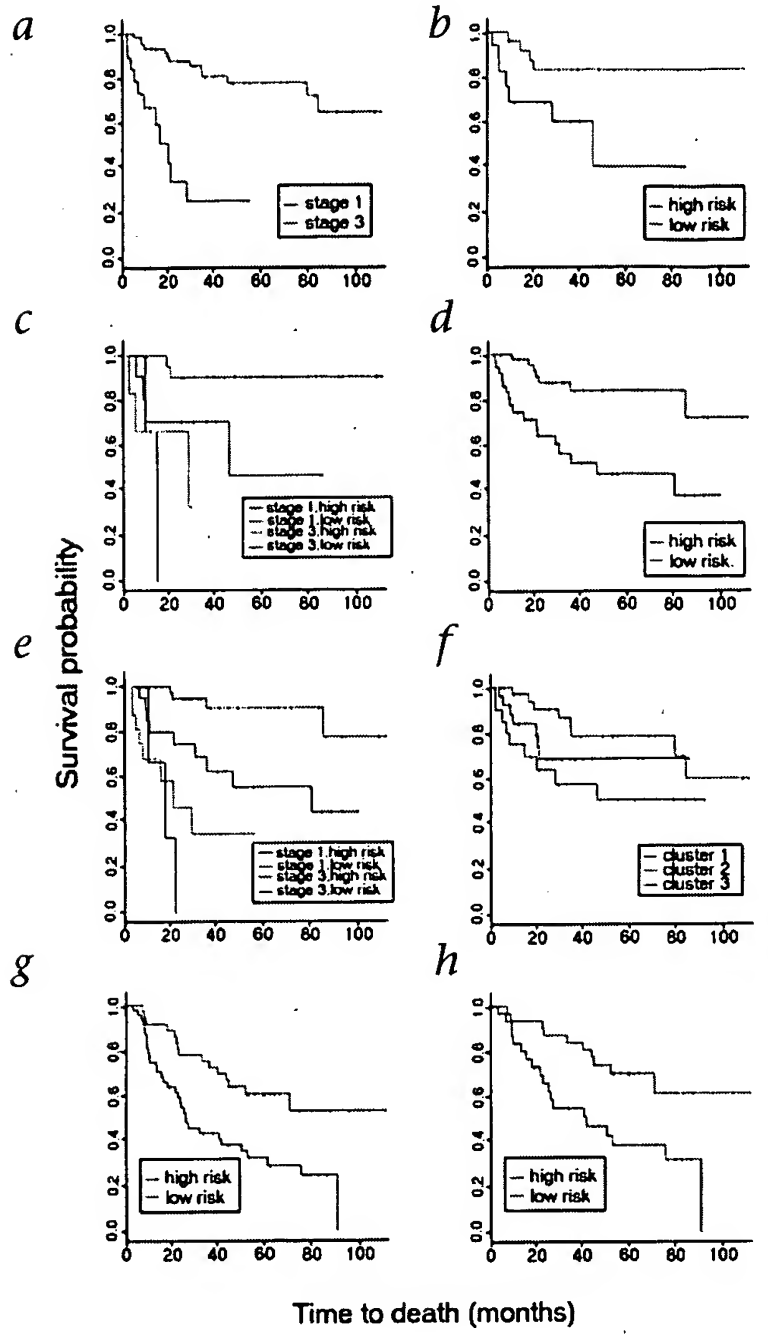


Fig. 3 Gene-expression profiles and patient survival. ***a***, Relationship between tumor stage and patient survival (stage 1 and stage 3 differ significantly, $P < 0.0001$). ***b***, Relationship between the survival in the 43 test samples and their risk assignments based on the 50-gene risk index estimated in the 43 training samples. The high- and low-risk groups differ significantly ($P = 0.024$). ***c***, Relationship between patient survival and the risk assignments in test samples (in *b)* conditional for tumor stage. The high- and low-risk stage I groups differ significantly ($P = 0.028$), whereas stage III low- and high-risk groups did not ($P = 0.634$). ***d***, Relationship between survival in the test cases and their risk assignments based on the 86 'leave-one-out' cross-validation of the 50-gene risk index. The high- and low-risk groups differ significantly ($P = 0.0006$). ***e***, Relationship between test case's risk assignment and survival (in *d)* conditional on tumor stage. The high- and low-risk stage I lung adenocarcinoma groups differ significantly from each other ($P = 0.003$), whereas low- and high-risk stage III tumors do not. ***f***, Relationship between tumor class identified by hierarchical clustering and patient survival. Survival for patients in Cluster 3 differed relative to the tumors in Cluster 2 ($P = 0.037$) and approached significance for Cluster 1 and 2 combined ($P = 0.06$). ***g***, Analysis of the Michigan-based risk index using top cross-validated survival genes identify a low- and high-risk group in an independent cohort of 84 Massachusetts-based lung adenocarcinomas that are significantly different ($P = 0.003$). ***h***, Among the 62 stage I lung adenocarcinomas in the Massachusetts sample, the high- and low-risk groups differed significantly ($P = 0.006$).

Immunoreactivity for both *IGFBP-3* and *HSP-70* (Fig. 2*c*) was detected in the cytoplasm of the adenocarcinomas, with little detectable reactivity in the stromal or inflammatory cells. Cystatin C was detected in alveolar pneumocytes and intra-alveolar macrophages in non-neoplastic lung parenchyma and also consistently in the cytoplasm of neoplastic cells.

Gene-expression profiles predict survival

As expected, Kaplan–Meier survival curves (Fig. 3*a*) and log-rank tests indicated poorer survival among stage III compared with stage I adenocarcinomas ($P = <0.0001$). Two statistical approaches were used to determine whether gene-expression profiles could predict survival using the data set of 4,966 genes. In one approach, equal numbers of randomly assigned stage I and stage III tumors constituted training ($n = 43$) and testing ($n = 43$) sets. In the training set, the top 10, 20, 50 or 75 genes were used to create risk indices that were evaluated for their association with survival using the 50th, 60th or 70th percentile cutoff points to categorize patients into high or low groups. The results were similar across cutoff points but the 50-gene risk index had the best overall association with survival in the training set. After conservatively choosing the 60th percentile cutoff point from the training set, we then applied this risk index and cutoff point to the testing set. The risk index of the top 50 genes correctly identified low- and high-risk individuals within the independent testing set ($P = 0.024$) (Fig. 3*b* and Supplementary Methods online). Notably, 11 stage I tumors were included in the high-risk subgroup. When this risk assignment was then conditionally examined for stage progression (Fig. 3*c*), low- and high-risk groups among stage I tumors were found to differ ($P = 0.028$) in their survival.

Identification of a robust set of survival genes

Although predictive of patient survival, a single training-testing set may not provide the most robust set of genes due to random sampling issues. Therefore, a 'leave-one-out' cross-validation approach was used to identify genes associated with survival from all 86-tumor samples. We first developed a 50-gene risk index in each training set, and then applied the risk index to the test case held out from the full set of tumors and assigned the held out tumor to the high- or low-risk groups (Fig. 3*d*). The high and low-risk subgroups determined in the test cases differed significantly in their overall survival ($P = 0.0006$). Among the larger group of stage I lung adenocarcinomas, the low-risk ($n = 46$) and high-risk ($n = 21$) groups had markedly different survival ($P = 0.003$) (Fig. 3*e*). Table 1 lists selected examples of the cumulative top 100 genes derived from this cross-validation procedure (complete list in Supplementary Table A online).

It was also noted that many of the stage I patients in the high-risk subgroup (Fig. 3*e*) were present in Cluster 3 (Fig. 1). Kaplan–Meier analysis (Fig. 3*f*) demonstrated a significantly worse survival ($P = 0.037$) for patients in Cluster 3 relative to patients in Cluster 2 and approaching significance for Cluster 1 and 2 combined ($P = 0.06$). This further indicates the important relationship between gene-expression profiles and patient survival, independent of disease stage.

Consistent with previous analyses of lung adenocarcinomas[23], 40% of stage I and 57.8% of stage III tumors had 12th or 13th codon K-*ras* gene mutations. Those patients with tumors containing K-*ras* mutations showed a trend of poorer survival, but

Table 1 Selected examples of the top 100 genes from cross-validation

Gene name	*P* (normal versus tumor t-test)	% Change in tumor	*P* (stage I versus stage III *t*-test)	% Change in stage III	Coefficient β	Unigene comment
						Apoptosis-related
CASP4	0.56	–6%	0.02	57%	0.0022	Caspase 4, apoptosis-related cysteine protease
P63	9.73E-04	37%	0.03	43%	0.0010	Transmembrane protein (63 kD), endoplasmic reticulum/Golgi intermediate compartment
						Cell adhesion and structure
KRT7	8.02E-08	126%	0.11	55%	0.0003	Keratin 7
LAMB1	0.14	–20%	0.01	60%	0.0027	Laminin, β1
						Cell cycle and growth regulators
BMP2	0.54	–21%	0.27	47%	0.0044	Bone morphogenetic protein 2
CDC6	1.31E-05	1070%	0.05	148%	0.0124	CDC6 (cell division cycle 6, *Saccharomyces cerevisiae* homolog)
S100P	2.10E-08	1572%	0.19	77%	0.0001	S100 calcium-binding protein P
SERPINE1	2.89E-03	72%	0.25	30%	0.0008	Serine (or cysteine) proteinase inhibitor, clade E (nexin)
STX1A	8.65E-08	54%	0.07	26%	0.0031	Syntaxin 1A (brain)
						Cell signaling
ADM	0.05	39%	0.04	117%	0.0016	adrenomedullin
AKAP 12	8.53E-03	-47%	0.05	214%	0.0010	A kinase (PRKA) anchor protein (gravin) 12
ARHE	0.06	–39%	0.05	87%	0.0092	ras homolog gene family, member E
GRB7	2.02E-03	38%	0.63	15%	0.0030	Growth factor receptor-bound protein 7
VEGF	6.50E-08	174%	0.02	85%	0.0013	Vascular endothelial growth factor
WNT10B	0.05	31%	0.48	20%	0.0022	Wingless-type MMTV integration site family, member 10B
						Chaperones
HSPA8	0.36	8%	9.01E-04	51%	0.0008	Heat-shock 70 kD protein 8
						Receptors
ERBB2	0.04	92%	0.37	120%	0.0013	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
FXYD3	0.10	111%	0.31	73%	0.0046	FXYD domain-containing ion transport regulator 3
SLC20A1	1.34E-03	58%	0.02	66%	0.0021	Solute carrier family 20 (phosphate transporter), member 1
						Enzymes, cellular metabolism
CSTB	1.57E-04	50%	0.15	34%	0.0001	Cystatin B (stefin B)
CTSL	0.48	–10%	0.03	67%	0.0007	Cathepsin L
CYP24	3.16E-06	N/A	0.97	2%	0.0008	Cytochrome P450, subfamily XXIV (vitamin D 24-hydroxylase)
FUT3	1.07E-07	114%	0.97	–1%	0.0033	Fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group included)
MLN64	0.20	32%	0.42	80%	0.0007	Steroidogenic acute regulatory protein related
PDE7A	0.12	33%	0.01	–35%	–0.0187	Phosphodiesterase 7A
PLGL	0.04	–68%	0.35	–170%	–0.0011	Plasminogen-like
SLC1A6	0.07	–32%	0.12	86%	0.0069	Solute carrier family 1 (high-affinity aspartate/glutamate transporter), member 6
						Transcription and translation
COPEB	0.10	–33%	0.26	25%	0.0016	Core promoter element binding protein
CRK	0.10	32%	0.03	48%	0.0098	v-crk avian sarcoma virus CT10 oncogene homolog
RELA	0.26	–7%	0.01	20%	0.0034	v-rel avian reticuloendotheliosis viral oncogene homolog A
						Unknown function
KIAA0005	2.21E-04	40%	0.02	45%	0.0010	KIAA0005 gene product
MGB1	0.27	125%	0.33	459%	0.0018	Mammaglobin 1

Bolded genes were also significant for survival in 43 tumor training set (Fig. 3*b*).

Table 1 Selected examples of the cumulative top 100 genes identified using training-testing, cross-validation of all 86 lung tumor samples. The percent change, as well as the direction, for the average values of the 10 non-neoplastic lung to all tumors, and for the 67 stage I to the 19 stage III tumors are shown. A positive coefficient β value is indicative of a relationship of gene expression to a poorer patient outcome. The genes are listed in potential functional categories. Genes that were also present in the top 50 survival genes using the 43-tumor training set (Fig. 3*b*) are indicated in bold type. Complete listing of the gene probe sets and annotated gene and unigene identifiers can be found in the Supplementary Methods.

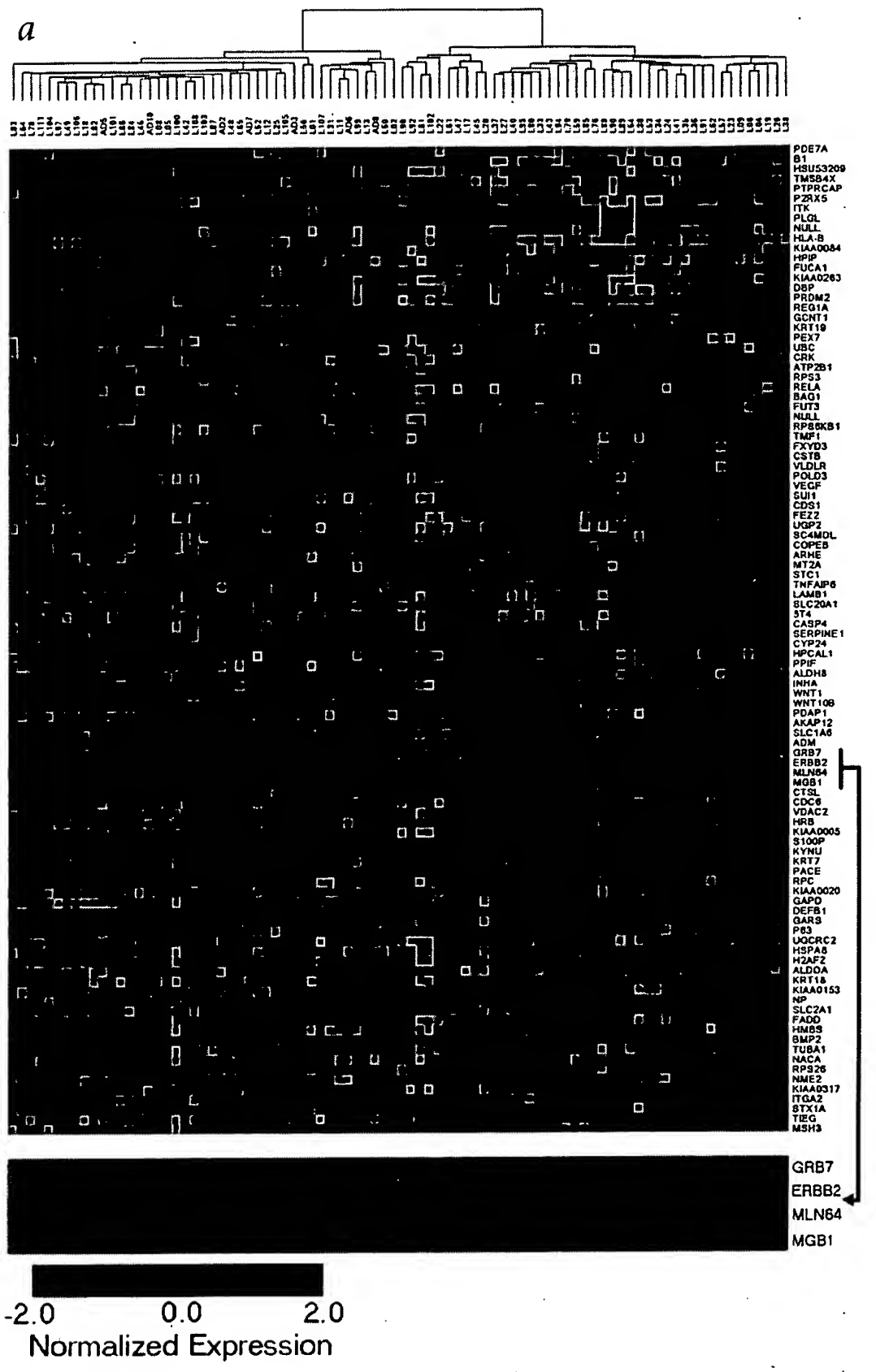


Fig. 4 Gene expression patterns of top survival genes ***a***, Gene-expression patterns determined using agglomerative hierarchical clustering of the 86 lung adenocarcinomas against the 100 survival-related genes (Table 1) identified by the training-testing, cross-validation analysis. Substantially elevated (red) or decreased (green) expression of the genes is observed in individual tumors. Some tumors (black arrow and expanded area) show extremely elevated expression of specific genes. ***b***, An outlier gene-expression pattern (>5 times the interquartile range among all samples) is observed for the *erb*B2 and *Reg*1A genes (top left and right, respectively). The *S100P* and *crk* genes (bottom left and right, respectively) show a graded pattern of expression related to patient survival. ○, alive; ●, dead (also in *c*). ***c***, The number of outliers per person identified in the top 100 genes plotted by survival distribution.

this difference did not reach statistical significance among all patients ($P = 0.25$), between patients within tumor clusters ($P = 0.41$) or when analyzed separately among stage I ($P = 0.22$) and stage III ($P = 0.53$) patients. Nuclear accumulation of p53 was detected in 17.9% stage I and in 22.2% stage III tumors. No significant relationship was observed for p53 staining and patient survival, cluster or tumor stage.

Confirmation using an independent set of adenocarcinomas

The robustness of our 50-gene risk index in predicting survival in lung adenocarcinomas was tested using oligonucleotide gene-expression data obtained from a completely independent (Massachusetts-based) sample of 84 lung adenocarcinomas (62 stage I, 14 stage II and 8 stage III; ref. 21, and dataset A at www.genome.wi.mit.edu/MPR/lung). To ensure equivalent power for testing and comparability of samples, the criteria for including tumors in the analysis were 40% or greater tumor cellularity, no mixed histology (that is, adenosquamous) and patient survival information. To obtain comparative gene-expression measures between the two data sets, gene sequences present on the U95A and HuGeneFL array were examined, and expression data for our top 50 cross-validation genes for all 84 Massachusetts samples were obtained and processed[24] (see also Supplementary Methods online). When we examined the risk assignment of these 84 samples, employing the identical cutoff point used for the 86 Michigan-based lung samples, we observed low- and high-risk groups (Fig. 3*g*; $P = 0.003$). Notably, among the 62 stage I tumors, high- and low-risk groups were observed that differed significantly ($P = 0.006$) in their survival (Fig. 3*h*).

Survival genes had graded and outlier expression patterns

A statistical and graphical analysis of the 100 survival-related genes (Table 1) clustered against all 86 tumors revealed individual tumors with substantially elevated expression in both a limited and larger number of genes (Fig. 4*a*). Among these genes, we observed two distinct patterns of expression related to patient survival. One pattern, designated 'outlier', included genes showing substantially elevated expression (greater than five times the interquartile range among all samples), whereas the other pattern, designated 'graded', was characterized by continuously distributed expression with patient survival (Fig. 4*b*). The *erb*B2 and *Reg*1A genes are examples of outlier expression patterns and *S100P* and *crk* genes of graded patterns. The number of outliers per person in the top 100 genes was identified and plotted according to survival times and events (Fig. 4*c*). Both stage I and stage III lung adenocarcinomas showed outlier gene patterns and 10 tumors contained 3 or more outlier genes.

Because gene amplification may result in increased gene expression, the nine genes with outlier expression patterns (*erb*B2, *SLC1*A6, *Wnt* 1, *MGB*1, *Reg*1A, *AKAP*12, *PACE*, *CYP*24, *KYNU*) and one gene with a graded expression pattern (*KRT*18) were examined using quantitative genomic PCR to evaluate genomic copy number (Fig. 5*a*). Gene amplification of *erb*B2 (17q12) was detected in tumor L94, which had the highest *erb*B2 mRNA expression (Fig. 4*a*). Gene amplification was not detected for any of the other seven tested genes in tumor L94, as well as in other tumors. The two genes most frequently demonstrating the outlier pattern in these lung adenocarcinomas were *KYNU* and *CYP*24, and were present in 10 and 9 tumors, respectively. *CYP*24 has been described as a gene amplified and overexpressed in breast cancer[25], and these results indicate elevated expression in lung adenocarcinoma.

To determine whether the graded or outlier gene-expression patterns also occur at the protein-expression level, 10 of the 100

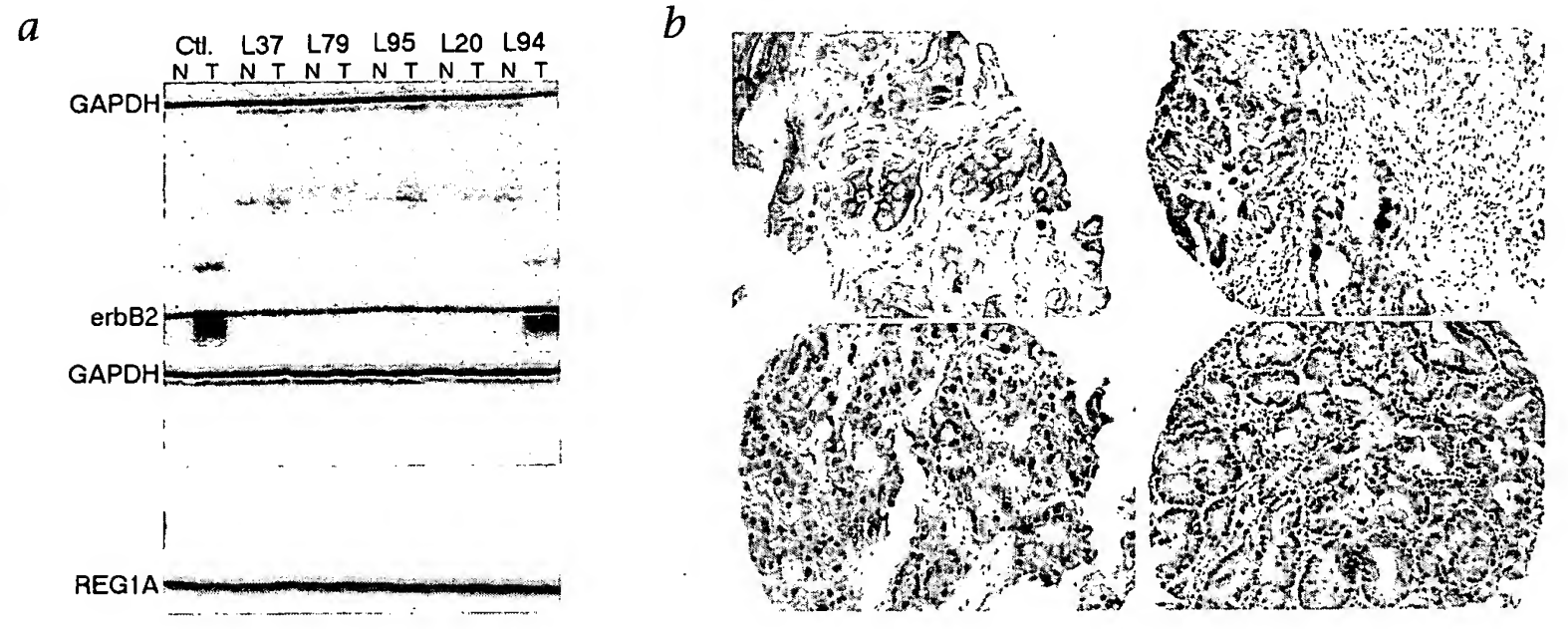


Fig. 5 Gene amplification and protein expression of survival-related genes. ***a***, Analysis of potential gene amplification for 9 genes showing outlier expression patterns in the lung tumors (*erb*B2, *SLC1*A6, *Wnt* 1, *MGB*1, *Reg*1A, *AKAP*12, *PACE*, *CYP*24 and *KYNU*) and examined using quantitative genomic PCR. A gene showing graded expression pattern (*KRT*18), and one gene (*PACE*4) with a similar chromosome location as *PACE*, were used as controls. Only *erb*B2 and *Reg*1A are shown. An esophageal adenocarcinoma with known high-level genomic amplification of *erb*B2 was used as a positive control and normal esophagus DNA was used as a negative control (Ctl). PCR fragments sizes were 343 bp for *GAPDH*, 166 bp for *erb*B2 and 126 bp for *Reg*1A. DNA is from normal lung (N) and tumor(T) from each patient (for example L37). ***b***, Immunohistochemical analysis of survival related genes with lung adenocarcinoma microarrays using the tumors from this study. The transmembrane erbB2 protein (top left) expression is substantially increased in tumor L94 containing the amplified *erb*B2 gene (Fig. 4*a* and *b*). Expression of VEGF (top right) and S100P (bottom left) was located within the neoplastic cells and the pattern of immunoreactivity was consistent with the graded expression pattern demonstrated by their mRNA profiles. Expression of the oncogene *crk* (bottom right) was abundantly expressed in neoplastic lung cells. Magnification, ×400 (erbB2); ×200 (VEGF, S100P and crk).

top survival genes (Table 1) for which specific antibodies were available were chosen for immunohistochemical analysis using lung-tumor arrays from this study (Fig. 5*b*). Expression of membrane erbB2 protein was substantially increased in the *erb*B2-amplified tumor L94 and very low levels of expression were present in other tumors, consistent with mRNA-expression measurements (Fig. 4*a* and *b*). CDC6 protein expression was also substantially higher in tumor L94, consistent with mRNA levels (data not shown). Expression of vascular endothelial growth factor (VEGF) and S100P (Fig. 5*b*), as well as cytokeratin 18 (KRT18), cytokeratin 7 (KRT7) and fas-associated death domain (FADD) protein (data not shown), was located within the lung tumor cells and consistent with the graded expression pattern of the mRNA profiles. The oncogene *crk* showed both graded mRNA as well as a graded protein-expression pattern with survival, and was abundantly expressed in the tumor cells (Fig. 5*b*). These results indicate that many survival-associated genes are expressed at the protein level and demonstrate similar mRNA and protein-expression patterns.

Discussion

We used several approaches for the analysis of gene-expression data related to clinicopathological variables and patient survival. One approach, hierarchical clustering, was used to examine similarities among lung adenocarcinomas in their patterns of gene expression. Previous studies of lung tumors[21,22] have also used this method to describe subclasses of lung tumors. Here, we found three clusters that showed significant differences with respect to tumor stage and tumor differentiation. This suggests, as expected, that tumors with similar histological features of differentiation demonstrate similarities in gene expression. This feature also partly underlies the observed statistical association of tumor stage and cluster, as many of the higher-stage tumors, often poorly differentiated and previously associated with a reduced survival[9,10], were located in Cluster 3. Although this cluster contained the highest percentage of stage III tumors, it also contained a nearly equal mixture of stage I and stage III tumors and not all tumors were poorly differentiated. This indicates that a subset of stage I lung adenocarcinomas share gene-expression profiles with higher-stage tumors. Notably, 10 of the 11 stage I tumors found in Cluster 3 were the high-risk stage I tumors identified using the risk index in the 'leave-one-out' cross-validation.

In contrast to previous analyses of lung adenocarcinomas[21,22], we validated the expression data from the arrays. The strong correlation of northern-blot analysis and oligonucleotide-array data for gene expression in the same samples (Fig. 2*b*) indicates that these studies provide robust gene-expression estimates. Immunohistochemistry using the same tumor samples in tissue arrays demonstrates protein expression within the lung tumor cells. Together, these studies indicate that many of the genes identified using gene-expression profiles are likely relevant to lung adenocarcinoma. For example, *IGFBP3* gene expression is increased in lung adenocarcinomas (Fig. 2*c*). IGFBP3 protein modulates the autocrine or paracrine effects of insulin-like growth factors, elevated IGFBP3 expression is observed in colon cancer[26], and increased serum IGFBP3 is associated with progression in breast cancer[27]. Heat-shock protein 70 (HSP-70) is increased in lung adenocarcinomas of smokers[28] and is associated with increased metastatic potential in breast cancer[29]. Increased serum lactate dehydrogenase is correlated with tumor stage and tumor burden[30], and cystatin C, a cysteine protease inhibitor expressed in human lung cancers[31], is prognostic in some cancers[32]. The decreased expression of this protease inhibitor may affect the invasive properties of the tumor cell.

The cross-validation analytical strategy we used is particularly informative for these types of gene-expression analyses for disease outcome[33,34], and identification of cross-validated genes with a larger tumor cohort may help refine this risk index for use in a clinical setting. The gene-expression data also provide opportunities to observe overarching patterns that advance our understanding of associations between genes and disease. For example, the top 100 survival genes include those involved in signaling, cell cycle and growth, transcription, translation and metabolism. Expression of many of these genes is likely a function of increased proliferation and metabolism in the more aggressive tumors. Some genes, such as *erb*B2 and *Reg*1A (Fig. 4*a* and *b*), were highly overexpressed in a few patients having poor survival. In one tumor, the *erb*B2 gene was amplified (Fig. 5*a*), demonstrating that genomic changes may underlie the overexpression of a subset of these outlier genes. Immunohistochemistry confirmed protein overexpression in this patient's tumor (Fig. 5*b*). Notably, seven of the eight outlier genes were not amplified, indicating that other mechanisms underlie the increased mRNA expression of these survival-related genes.

Most genes showed a graded relationship between expression and patient survival. Genes such as that encoding VEGF, known to be strongly associated with survival in lung cancer[35,36] were identified as related to patient survival in our study. *VEGF* demonstrated a graded expression pattern, as did the S100P and *crk* oncogene (Fig. 5*b*). S100P is a calcium-regulated protein not previously reported in lung cancer. The *crk* gene, the cellular homolog of the v-*crk* oncogene, is a member of a family of adaptor proteins involved in signal transduction and interacts directly with c-jun N-terminal kinase 1 (JNK1)[37]. Although *crk* has not been shown to have a role lung cancer, its role in the MAP-kinase pathway, which leads to activation of matrix metalloproteinase secretion and cell invasion[38], indicates potential involvement in the the tumor cell invasion or metastasis of some lung adenocarcinomas. Among the many genes identified in this study, like *crk*, that may be causally involved in lung cancer progression (Table 1), some were related to survival in many patients, and others in only smaller subsets of patients. This result is consistent with the complex molecular architecture of tumors in general, the heterogeneity of lung adenocarcinomas in particular and the multiple mechanisms underlying tumor-cell survival, invasion and metastasis[39].

Our results demonstrate that a gene-expression risk profile—based on the genes most associated with patient survival—can distinguish stage I lung adenocarcinomas and differentiate prognoses. The particular genes that define the clusters, or are associated with survival, likely reflect the characteristics of the particular tumors included in the analysis. Current therapy for patients with stage I disease usually consists of surgical resection without adjuvant treatment[2,3]. Clearly, the identification of a high-risk group among patients with stage I disease would lead to consideration of additional therapeutic intervention for this group, possibly leading to improved survival of these patients.

Methods

Patient population. Sequential patients seen at the University of Michigan Hospital between May 1994 and July 2000 for stage I or stage III lung adenocarcinoma were evaluated for this study. Consent was received and the project was approved by the local Institutional Review Board. Primary tumors and adjacent non-neoplastic lung tissue were obtained at the time of

surgery. Peripheral portions of resected lung carcinomas were sectioned, evaluated by a study pathologist and compared with routine H&E sections of the same tumors, and utilized for mRNA isolation. Regions chosen for analysis contained a tumor cellularity greater than 70%, no mixed histology, potential metastatic origin, extensive lymphocytic infiltration or fibrosis. Tumors were histopathologically divided into two categories based on their growth pattern: bronchial-derived, if they exhibited invasive features with architectural destruction, and bronchioloalveolar, if they exhibited preservation of the lung architecture. All stage I patients received only surgical resection with intra-thoracic nodal sampling and no other treatments. Stage III patients received surgical resection plus chemotherapy and radiotherapy.

Gene-expression profiling and K-*ras* mutation analysis. RNA isolation, cRNA synthesis and gene-expression profiling were performed as described[24]. Details of gene annotation and K-*ras* mutation analysis are provided in supplementary information.

Northern-blot analysis. Total cellular RNA (10 μg) was separated in 1.2% agarose-formaldehyde gels and vacuum-transferred to Gene Screen Plus (NEN Life Science Products, Boston, Massachusetts). Hybridization conditions and probe labeling were as described[40]. Individual sequence-validated cDNA image clones for human *IGFBP3* (clone 1407750), *LDH-A* (clone 2420241), cystatin C (CTS3; clone 949938) were from Research Genetics (Huntsville, Alabama). The human histone H4 cDNA and the *28S* ribosomal RNA 26-mer oligonucleotide probe were prepared and labeled as described[40].

Gene-amplification analysis. 11 genes were selected for the analysis of genomic alterations. Primers were designed using PrimerSelect 4.05 Windows 32 software (DNASTAR, Madison, Wisconsin), avoiding pseudogenes or potential homologous regions. Forward and reverse primers for the genes are provided (Supplementary Methods online). Quantitative genomic-PCR was then applied and analyzed as described[41].

Immunohistochemical staining. The H&E-stained slides of all primary lung tumors were used to identify the most representative regions of each tumor and a tissue microarray (TMA) block was constructed as described[42]. Immunohistochemistry (IHC) was performed using both routine and sections from the TMA block as described[24]. Detailed methods and the concentrations used for all antibodies are provided in the Supplementary Methods.

Statistical methods. *t*-tests were used to identify differences in mean gene-expression levels between comparison groups. Agglomerative hierarchical clustering[43] was applied using the average linkage method to investigate whether there was evidence for natural groupings of tumor samples based on correlations between gene-expression profiles. To investigate the robustness of the clustering inference, gene-expression values were perturbed by adding random Gaussian error of magnitude obtained from a duplicate sample to each data point and then reclustered to determine concordance in the tumor's class membership. Pearson, χ^2 and Fisher's exact tests were used to assess whether cluster membership was associated with physical and genetic characteristics of the tumors.

To determine whether gene-expression profiles were associated with variability in survival times, 2 separate but complementary approaches were used. In the first approach, the 86 tumors were randomly assigned to equivalent training and testing sets consisting of equal numbers of stage I and III tumors in order to validate a novel risk-index function that captured the effect of many genes at once. In the second approach, cross-validation[44] was used to more robustly identify the genes associated with survival. Briefly, a 'leave-one-out' cross-validation procedure in which 85 of the 86 tumors (the training set) was used to identify genes that were univariately associated with survival. The risk index was defined as a linear combination of the gene-expression values for the top genes identified by univariate Cox proportional-hazard regression modeling[45], weighted by their estimated regression coefficients. Kaplan–Meier survival plots and log-rank tests were then used to assess whether the risk-index assignment to high/low categories was validated in the test set. A more detailed description is provided (Supplementary Methods online).

Note: Supplementary information is available on the Nature Medicine website.

Acknowledgments

We thank D. Sanders for technical assistance; D. Sing for assistance with the figures; and G. Omenn for critical reading of this manuscript. This work was supported by National Cancer Institute grant: U19 CA-85953 and the Tissue Core of the University of Michigan Comprehensive Cancer Center (NIH CA-46952).

Competing interests statement

The authors declare that they have no competing financial interests.

RECEIVED 5 APRIL; ACCEPTED 14 JUNE 2002

1. Fry, W.A., Phillips, J.L. & Menck, H.R. Ten-year survey of lung cancer treatments and survival in hospitals in the United States. *Cancer* **86**, 1867–1876 (1999).
2. Williams, D.E. *et al.* Survival of patients surgically treated for stage I lung cancer. *J. Thorac. Cardiovasc. Surg.* **82**, 70–76 (1981).
3. Pairolero, P.C. *et al.* Postsurgical stage I bronchogenic carcinoma: Morbid implications of recurrent disease. *Ann. Thorac. Surg.* **38**, 331–338 (1984).
4. Naruke, T. *et al.* Prognosis and survival in resected carcinoma based on the new international staging system. *J. Thorac. Cardiovasc. Surg.* **96**, 440–447 (1988).
5. Kaisermann, M.C. *et al.* Evolving features of lung adenocarcinoma in Rio de Janeiro, Brazil. *Oncol. Rep.* **8**, 189–192 (2001).
6. Roggli, V.L. *et al.* Lung cancer heterogeneity: A blinded and randomized study of 100 consecutive cases. *Hum. Pathol.* **16**, 569–579 (1985).
7. Gail, M.H. *et al.* Prognostic factors in patients with resected stage I non-small cell lung cancer: A report from the Lung Cancer Study Group. *Cancer* **54**, 1802–1813 (1984).
8. Takise, A. *et al.* Histopathologic prognostic factors in adenocarcinomas of the peripheral lung less than 2 cm in diameter. *Cancer* **61**, 2083–2088 (1988).
9. Ichinose, Y. *et al.* Is T factor of the TMN staging system a predominant prognostic factor in pathologic stage I non-small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* **106**, 90–94 (1993).
10. Harpole, D.H. *et al.* A prognostic model of recurrence and death in stage I non-small cell lung cancer utilizing presentation, histopathology, and oncoprotein expression. *Cancer Res.* **55**, 51–56 (1995).
11. Rodenhuis, S. *et al.* Mutational activation of the K-*ras* oncogene: A possible pathogenic factor in adenocarcinoma of the lung. *N. Engl. J. Med.* **317**, 929–935 (1987).
12. Slebos, R.J.C. *et al.* K-*ras* oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N. Engl. J. Med.* **323**, 561–565 (1990).
13. Horio, Y. *et al.* Prognostic significance of p53 mutations and 3p deletions in primary resected non-small cell lung cancer. *Cancer Res.* **53**, 1–4 (1993).
14. Kern, J.A. *et al.* C-*erb*B-2 expression and codon 12 K-*ras* mutations both predict shortened survival for patients with pulmonary adenocarcinomas. *J. Clin. Invest.* **93**, 516–520 (1994).
15. Ebina, M. *et al.* Relationship of p53 overexpresson and up-regulation of proliferating cell nuclear antigen with the clinical course of non-small cell lung cancer. *Cancer Res.* **54**, 2496–2503 (1994).
16. Mehdi, S.A. *et al.* Prognostic markers in resected stage I and II non-small cell lung cancer: an analysis of 260 patients with 5 year follow-up. *Clin. Lung Cancer* **1**, 59–67 (1997).
17. Schneider, P.M. *et al.* Multiple molecular marker testing (p53, c-Ki-*ras*, c-erbB-2) improves estimation of prognosis in potentially curative resected non-small cell lung cancer. *Br. J. Cancer* **83**, 473–479 (2000).
18. Herbst, R.S. *et al.* Differential expression of E-cadherin and type IV collagenase genes predicts outcome in patients with stage I non-small cell lung carcinoma. *Clin. Can. Res.* **6**, 790–797 (2000).
19. Liotta, L. & Petricion, E. Molecular profiling of human cancer. *Nature Rev. Genet.* **1**, 48–56 (2000).
20. Golub, T.R. Editorial: Genome-wide views of cancer. *N. Engl. J. Med.* **344**, 601–602 (2001).
21. Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98**, 13790–13795 (2001).
22. Garber, M.E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* **98**, 13784–13789 (2001).
23. Mills, N.E. *et al.* Increased prevalence of K-*ras* oncogene mutations in lung adenocarcinoma. *Cancer Res.* **55**, 1444–1447 (1995).
24. Giordano T.J. *et al.* Organ-specific molecular classification of lung, colon and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.* **159**, 1231–1238 (2001).
25. Albertson, D.G. *et al.* Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nature Genet.* **25**, 144–146 (2000).
26. Kansra, S. *et al.* IGFBP-3 mediates TGF β1 proliferative response in colon cancer cells. *Int. J. Cancer* **87**, 373–378 (2000).
27. Vadgama J.V. *et al.* Plasma insulin-like growth factor-I and serum IGF-binding protein 3 can be associated with the progression of breast cancer, and predict the risk of recurrence and the probability of survival in African-American and Hispanic

women. *Oncology* 57, 330–340 (1999).
28. Volm, M., Mattern, J. & Stammler, G. Up-regulation of heat shock protein 70 in adenocarcinoma of the lung in smokers. *Anticancer Res.* 15, 2607–2609 (1995).
29. Ciocca, D.R. *et al.* Heat shock protein hsp70 in patients with auxillary lymph node-positive breast cancer:prognostic implications. *J. Natl. Cancer. Inst.* 85, 570–574 (1993).
30. Rotenberg, Z. *et al.* Total lactate dehydrogenase and its isoenzymes in serum of patients with non-small cell lung cancer. *Clin. Chem.* 34, 668–670 (1988).
31. Krepela, E. *et al.* Cysteine proteases and cysteine protease inhibitors in non-small cell lung cancer. *Neoplasma* 45, 318–331 (1998).
32. Kos, J. *et al.* Cysteine proteinases and their inhibitors in extracellular fluids: Markers for diagnosis and prognosis in cancer. *Int. J. Biol. Markers* 15, 84–89 (2000).
33. Golub, T.R. *et al.* Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537 (1999).
34. Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* 344, 539–548 (2001).
35. Ohta, Y. *et al.* Vascular endothelial growth factor and lymph node metastasis in primary lung cancer. *Br. J. Cancer.* 76, 1041–1045 (1997).
36. Shibusa, T., Shijubo, N. & Abe, S. Tumor angiogenesis and vascular endothelial growth factor expression in stage I lung adenocarcinoma. *Clin. Cancer Res.* 4, 1483–1487 (1998).
37. Girardin, S.E. & Yaniv, M. A direct interaction between JNK1 and CrkII is critical for Rac1-induced JNK activation. *EMBO J.* 20, 3437–3446 (2001).
38. Liu, E. *et al.* The Ras-mitogen-activated protein kinase pathway is critical for the activation of matrix metalloproteinase secretion and the invasiveness in v-crk-transformed 3Y1. *Cancer Res.* 60, 2361–64 (2000).
39. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* 100, 57–70 (2000).
40. Hanson, L.A. *et al.* Expression of the glucocorticoid receptor and K-*ras* genes in urethan-induced mouse lung tumors and transformed cell lines. *Exp. Lung. Res.* 17, 371–387 (1991).
41. Lin, L. *et al.* A minimal critical region of the 8p22-23 amplicon in esophageal adenocarcinomas defined using STS-amplification mapping and quantitative PCR includes the GATA-4 gene. *Cancer Res.* 60, 1341–1347 (2000).
42. Kononen, J. *et al.* Tissue microarrays for high throughput molecular profiling of tumor specimens. *Nature Med.* 4, 844–847 (1998).
43. Johnson, R. & Wichern, D.W. *Applied Multivariate Statistical Analysis.* 543–578 (Prentice Hall, New Jersey, 1988).
44. Stone, M. Asymptomics for and against cross-validation. *Biometrika* 64, 29–38 (1977).
45. Cox, D.R. Regression models and life tables. *J.R. Stat. Soc.* 34, 187–220 (1972).